LNAI 5638

Dylan D. Schmorrow
Ivy V. Estabrooke
Marc Grootjen (Eds.)

Foundations of Augmented Cognition

Neuroergonomics and Operational Neuroscience

5th International Conference, FAC 2009
Held as Part of HCI International 2009
San Diego, CA, USA, July 2009, Proceedings

HCI2009
INTERNATIONAL

Springer

Lecture Notes in Artificial Intelligence 5638

Edited by R. Goebel, J. Siekmann, and W. Wahlster

Subseries of Lecture Notes in Computer Science

Dylan D. Schmorrow   Ivy V. Estabrooke
Marc Grootjen (Eds.)

Foundations of Augmented Cognition

Neuroergonomics and Operational Neuroscience

5th International Conference, FAC 2009
Held as Part of HCI International 2009
San Diego, CA, USA, July 19-24, 2009
Proceedings

Springer

Series Editors

Randy Goebel, University of Alberta, Edmonton, Canada
Jörg Siekmann, University of Saarland, Saarbrücken, Germany
Wolfgang Wahlster, DFKI and University of Saarland, Saarbrücken, Germany

Volume Editors

Dylan D. Schmorrow
Ivy V. Estabrooke
United States Navy
Office of Naval Research
875 North Randolph Road, Arlington, VA 22203, USA
E-mail: schmorrow@yahoo.com, ivy.estabrooke@navy.mil

Marc Grootjen
Royal Netherlands Navy
Defense Materiel Organization
P.O. Box 20702, 2500 ES The Hague, The Netherlands
E-mail: marc@grootjen.nl

Library of Congress Control Number: Applied for

CR Subject Classification (1998): I.2, J.3, I.4, H.3, C.2

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN	0302-9743
ISBN-10	3-642-02811-X Springer Berlin Heidelberg New York
ISBN-13	978-3-642-02811-3 Springer Berlin Heidelberg New York

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12712670 06/3180 5 4 3 2 1 0

Foreword

The 13th International Conference on Human–Computer Interaction, HCI International 2009, was held in San Diego, California, USA, July 19–24, 2009, jointly with the Symposium on Human Interface (Japan) 2009, the 8th International Conference on Engineering Psychology and Cognitive Ergonomics, the 5th International Conference on Universal Access in Human–Computer Interaction, the Third International Conference on Virtual and Mixed Reality, the Third International Conference on Internationalization, Design and Global Development, the Third International Conference on Online Communities and Social Computing, the 5th International Conference on Augmented Cognition, the Second International Conference on Digital Human Modeling, and the First International Conference on Human Centered Design.

A total of 4,348 individuals from academia, research institutes, industry and governmental agencies from 73 countries submitted contributions, and 1,397 papers that were judged to be of high scientific quality were included in the program. These papers address the latest research and development efforts and highlight the human aspects of the design and use of computing systems. The papers accepted for presentation thoroughly cover the entire field of human–computer interaction, addressing major advances in knowledge and effective use of computers in a variety of application areas.

This volume, edited by Dylan Schmorrow, Ivy Estabrooke, and Marc Grootjen, contains papers in the thematic area of Augmented Cognition, addressing the following major topics:

- Understanding Human Cognition and Behavior in Complex Tasks and Environments
- Cognitive Modeling, Perception, Emotion and Interaction
- Cognitive Load and Performance
- Electroencephalography and Brain Activity Measurement
- Physiological Measuring
- Augmented Cognition in Training and Education
- Brain-Computer Interfaces
- Rehabilitation and Cognitive Aids

The remaining volumes of the HCI International 2009 proceedings are:

- Volume 1, LNCS 5610, Human–Computer Interaction—New Trends (Part I), edited by Julie A. Jacko
- Volume 2, LNCS 5611, Human–Computer Interaction—Novel Interaction Methods and Techniques (Part II), edited by Julie A. Jacko
- Volume 3, LNCS 5612, Human–Computer Interaction—Ambient, Ubiquitous and Intelligent Interaction (Part III), edited by Julie A. Jacko
- Volume 4, LNCS 5613, Human–Computer Interaction—Interacting in Various Application Domains (Part IV), edited by Julie A. Jacko

- Volume 5, LNCS 5614, Universal Access in Human–Computer Interaction—Addressing Diversity (Part I), edited by Constantine Stephanidis
- Volume 6, LNCS 5615, Universal Access in Human–Computer Interaction—Intelligent and Ubiquitous Interaction Environments (Part II), edited by Constantine Stephanidis
- Volume 7, LNCS 5616, Universal Access in Human–Computer Interaction—Applications and Services (Part III), edited by Constantine Stephanidis
- Volume 8, LNCS 5617, Human Interface and the Management of Information—Designing Information Environments (Part I), edited by Michael J. Smith and Gavriel Salvendy
- Volume 9, LNCS 5618, Human Interface and the Management of Information—Information and Interaction (Part II), edited by Gavriel Salvendy and Michael J. Smith
- Volume 10, LNCS 5619, Human Centered Design, edited by Masaaki Kurosu
- Volume 11, LNCS 5620, Digital Human Modeling, edited by Vincent G. Duffy
- Volume 12, LNCS 5621, Online Communities and Social Computing, edited by A. Ant Ozok and Panayiotis Zaphiris
- Volume 13, LNCS 5622, Virtual and Mixed Reality, edited by Randall Shumaker
- Volume 14, LNCS 5623, Internationalization, Design and Global Development, edited by Nuray Aykin
- Volume 15, LNCS 5624, Ergonomics and Health Aspects of Work with Computers, edited by Ben-Tzion Karsh
- Volume 17, LNAI 5639, Engineering Psychology and Cognitive Ergonomics, edited by Don Harris

I would like to thank the Program Chairs and the members of the Program Boards of all thematic areas, listed below, for their contribution to the highest scientific quality and the overall success of HCI International 2009.

Ergonomics and Health Aspects of Work with Computers

Program Chair: Ben-Tzion Karsh

Human Interface and the Management of Information

Program Chair: Michael J. Smith

Gunilla Bradley, Sweden
Hans-Jörg Bullinger, Germany
Alan Chan, Hong Kong
Klaus-Peter Fähnrich, Germany
Michitaka Hirose, Japan
Jhilmil Jain, USA
Yasufumi Kume, Japan
Mark Lehto, USA
Fiona Fui-Hoon Nah, USA
Shogo Nishida, Japan
Robert Proctor, USA
Youngho Rhee, Korea
Anxo Cereijo Roibás, UK
Katsunori Shimohara, Japan
Dieter Spath, Germany
Tsutomu Tabe, Japan
Alvaro D. Taveira, USA
Kim-Phuong L. Vu, USA
Tomio Watanabe, Japan
Sakae Yamamoto, Japan
Hidekazu Yoshikawa, Japan
Li Zheng, P.R. China
Bernhard Zimolong, Germany

Human–Computer Interaction

Program Chair: Julie A. Jacko

Sebastiano Bagnara, Italy
Sherry Y. Chen, UK
Marvin J. Dainoff, USA
Jianming Dong, USA
John Eklund, Australia
Xiaowen Fang, USA
Ayse Gurses, USA
Vicki L. Hanson, UK
Sheue-Ling Hwang, Taiwan
Wonil Hwang, Korea
Yong Gu Ji, Korea
Steven Landry, USA
Gitte Lindgaard, Canada
Chen Ling, USA
Yan Liu, USA
Chang S. Nam, USA
Celestine A. Ntuen, USA
Philippe Palanque, France
P.L. Patrick Rau, P.R. China
Ling Rothrock, USA
Guangfeng Song, USA
Steffen Staab, Germany
Wan Chul Yoon, Korea
Wenli Zhu, P.R. China

Engineering Psychology and Cognitive Ergonomics

Program Chair: Don Harris

Guy A. Boy, USA
John Huddlestone, UK
Kenji Itoh, Japan
Hung-Sying Jing, Taiwan
Ron Laughery, USA
Wen-Chin Li, Taiwan
James T. Luxhøj, USA
Nicolas Marmaras, Greece
Sundaram Narayanan, USA
Mark A. Neerincx, The Netherlands
Jan M. Noyes, UK
Kjell Ohlsson, Sweden
Axel Schulte, Germany
Sarah C. Sharples, UK

Neville A. Stanton, UK
Xianghong Sun, P.R. China
Andrew Thatcher, South Africa
Matthew J.W. Thomas, Australia
Mark Young, UK

Universal Access in Human–Computer Interaction

Program Chair: Constantine Stephanidis

Julio Abascal, Spain
Ray Adams, UK
Elisabeth André, Germany
Margherita Antona, Greece
Chieko Asakawa, Japan
Christian Bühler, Germany
Noelle Carbonell, France
Jerzy Charytonowicz, Poland
Pier Luigi Emiliani, Italy
Michael Fairhurst, UK
Dimitris Grammenos, Greece
Andreas Holzinger, Austria
Arthur I. Karshmer, USA
Simeon Keates, Denmark
Georgios Kouroupetroglou, Greece
Sri Kurniawan, USA
Patrick M. Langdon, UK
Seongil Lee, Korea
Zhengjie Liu, P.R. China
Klaus Miesenberger, Austria
Helen Petrie, UK
Michael Pieper, Germany
Anthony Savidis, Greece
Andrew Sears, USA
Christian Stary, Austria
Hirotada Ueda, Japan
Jean Vanderdonckt, Belgium
Gregg C. Vanderheiden, USA
Gerhard Weber, Germany
Harald Weber, Germany
Toshiki Yamaoka, Japan
Panayiotis Zaphiris, UK

Virtual and Mixed Reality

Program Chair: Randall Shumaker

Pat Banerjee, USA
Mark Billinghurst, New Zealand
Charles E. Hughes, USA
David Kaber, USA
Hirokazu Kato, Japan
Robert S. Kennedy, USA
Young J. Kim, Korea
Ben Lawson, USA
Gordon M. Mair, UK
Miguel A. Otaduy, Switzerland
David Pratt, UK
Albert “Skip” Rizzo, USA
Lawrence Rosenblum, USA
Dieter Schmalstieg, Austria
Dylan Schmorrow, USA
Mark Wiederhold, USA

Internationalization, Design and Global Development

Program Chair: Nuray Aykin

Michael L. Best, USA
Ram Bishu, USA
Alan Chan, Hong Kong
Andy M. Dearden, UK
Susan M. Dray, USA
Vanessa Evers, The Netherlands
Paul Fu, USA
Emilie Gould, USA

Sung H. Han, Korea
Veikko Ikonen, Finland
Esin Kiris, USA
Masaaki Kurosu, Japan
Apala Lahiri Chavan, USA
James R. Lewis, USA
Ann Light, UK
James J.W. Lin, USA
Rungtai Lin, Taiwan
Zhengjie Liu, P.R. China
Aaron Marcus, USA
Allen E. Milewski, USA
Elizabeth D. Mynatt, USA
Oguzhan Ozcan, Turkey
Girish Prabhu, India
Kerstin Röse, Germany
Eunice Ratna Sari, Indonesia
Supriya Singh, Australia
Christian Sturm, Spain
Adi Tedjasaputra, Singapore
Kentaro Toyama, India
Alvin W. Yeo, Malaysia
Chen Zhao, P.R. China
Wei Zhou, P.R. China

Online Communities and Social Computing

Program Chairs: A. Ant Ozok, Panayiotis Zaphiris

Chadia N. Abras, USA
Chee Siang Ang, UK
Amy Bruckman, USA
Peter Day, UK
Fiorella De Cindio, Italy
Michael Gurstein, Canada
Tom Horan, USA
Anita Komlodi, USA
Piet A.M. Kommers, The Netherlands
Jonathan Lazar, USA
Stefanie Lindstaedt, Austria
Gabriele Meiselwitz, USA
Hideyuki Nakanishi, Japan
Anthony F. Norcio, USA
Jennifer Preece, USA
Elaine M. Raybourn, USA
Douglas Schuler, USA
Gilson Schwartz, Brazil
Sergei Stafeev, Russia
Charalambos Vrasidas, Cyprus
Cheng-Yen Wang, Taiwan

Augmented Cognition

Program Chair: Dylan D. Schmorrow

Andy Bellenkes, USA
Andrew Belyavin, UK
Joseph Cohn, USA
Martha E. Crosby, USA
Tjerk de Greef, The Netherlands
Blair Dickson, UK
Traci Downs, USA
Julie Drexler, USA
Ivy Estabrooke, USA
Cali Fidopiastis, USA
Chris Forsythe, USA
Wai Tat Fu, USA
Henry Girolamo, USA
Marc Grootjen, The Netherlands
Taro Kanno, Japan
Wilhelm E. Kincses, Germany
David Kobus, USA
Santosh Mathan, USA
Rob Matthews, Australia
Dennis McBride, USA
Robert McCann, USA
Jeff Morrison, USA
Eric Muth, USA
Mark A. Neerincx, The Netherlands
Denise Nicholson, USA
Glenn Osga, USA

Dennis Proffitt, USA
Leah Reeves, USA
Mike Russo, USA
Kay Stanney, USA
Roy Stripling, USA
Mike Swetnam, USA
Rob Taylor, UK
Maria L. Thomas, USA
Peter-Paul van Maanen, The Netherlands
Karl van Orden, USA
Roman Vilimek, Germany
Glenn Wilson, USA
Thorsten Zander, Germany

Digital Human Modeling

Program Chair: Vincent G. Duffy

Karim Abdel-Malek, USA
Thomas J. Armstrong, USA
Norm Badler, USA
Kathryn Cormican, Ireland
Afzal Godil, USA
Ravindra Goonetilleke, Hong Kong
Anand Gramopadhye, USA
Sung H. Han, Korea
Lars Hanson, Sweden
Pheng Ann Heng, Hong Kong
Tianzi Jiang, P.R. China
Kang Li, USA
Zhizhong Li, P.R. China
Timo J. Määttä, Finland
Woojin Park, USA
Matthew Parkinson, USA
Jim Potvin, Canada
Rajesh Subramanian, USA
Xuguang Wang, France
John F. Wiechel, USA
Jingzhou (James) Yang, USA
Xiu-gan Yuan, P.R. China

Human Centered Design

Program Chair: Masaaki Kurosu

Gerhard Fischer, USA
Tom Gross, Germany
Naotake Hirasawa, Japan
Yasuhiro Horibe, Japan
Minna Isomursu, Finland
Mitsuhiko Karashima, Japan
Tadashi Kobayashi, Japan
Kun-Pyo Lee, Korea
Loïc Martínez-Normand, Spain
Dominique L. Scapin, France
Haruhiko Urokohara, Japan
Gerrit C. van der Veer, The Netherlands
Kazuhiko Yamazaki, Japan

In addition to the members of the Program Boards above, I also wish to thank the following volunteer external reviewers: Gavin Lew from the USA, Daniel Su from the UK, and Ilia Adami, Ioannis Basdekis, Yannis Georgalis, Panagiotis Karampelas, Iosif Klironomos, Alexandros Mourouzis, and Stavroula Ntoa from Greece.

This conference could not have been possible without the continuous support and advice of the Conference Scientific Advisor, Prof. Gavriel Salvendy, as well as the dedicated work and outstanding efforts of the Communications Chair and Editor of HCI International News, Abbas Moallem.

I would also like to thank for their contribution toward the organization of the HCI International 2009 conference the members of the Human–Computer Interaction Laboratory of ICS-FORTH, and in particular Margherita Antona, George Paparoulis, Maria Pitsoulaki, Stavroula Ntoa, and Maria Bouhli.

Constantine Stephanidis

HCI International 2011

The 14th International Conference on Human–Computer Interaction, HCI International 2011, will be held jointly with the affiliated conferences in the summer of 2011. It will cover a broad spectrum of themes related to human–computer interaction, including theoretical issues, methods, tools, processes and case studies in HCI design, as well as novel interaction techniques, interfaces and applications. The proceedings will be published by Springer. More information about the topics, as well as the venue and dates of the conference, will be announced through the HCI International Conference series website: http://www.hci-international.org/

General Chair
Professor Constantine Stephanidis
University of Crete and ICS-FORTH
Heraklion, Crete, Greece
Email: cs@ics.forth.gr

Table of Contents

Part I: Understanding Human Cognition and Behavior in Complex Tasks and Environments

Part II: Cognitive Modeling, Perception, Emotion and Interaction

Part VI: Augmented Cognition in Training and Education

Part VII: Brain-Computer Interfaces

Part VIII: Rehabilitation and Cognitive Aids

A Generic Personal Assistant Agent Model for Support in Demanding Tasks

Tibor Bosse[1], Rob Duell[2], Mark Hoogendoorn[1], Michel Klein[1], Rianne van Lambalgen[1], Andy van der Mee[2], Rogier Oorburg[2], Alexei Sharpanskykh[1], Jan Treur[1], and Michael de Vos[2]

[1] Vrije Universiteit Amsterdam, Department of Artificial Intelligence
De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands
{tbosse,mhoogen,mcaklein,rm.van.lambalgen,sharp,treur}@few.vu.nl
[2] Force Vision Lab, Barbara Strozzilaan 362a, 1083 HN Amsterdam, The Netherlands
{rob,andy,rogier,michael}@forcevisionlab.nl

Abstract. Human task performance may vary depending on the characteristics of the human, the task and the environment over time. To ensure high effectiveness and efficiency of the execution of tasks, automated personal assistance may be provided to task performers. A personal assistant agent may constantly monitor the human's state and task execution, analyse the state of the human and task, and intervene when a problem is detected. This paper proposes a generic design for a Personal Assistant agent model which can be deployed in a variety of domains. Application of the Personal Assistant model is illustrated by a case study from the naval domain.

1 Introduction

Human task performance can degrade over time when demanding tasks are being performed. Such degradation can for instance be caused by available resources being exceeded [1]. Furthermore, the effectiveness and efficiency of the task execution are often dependent on the capabilities, experience, and condition of the actor performing the task. Different actors may require different degrees of assistance and various resources for the task execution. High effectiveness and efficiency levels are of particular importance for critical tasks. Furthermore, as a longer term aim, the human should remain healthy during the processes of task execution. To overcome the limitations of human cognition (e.g. in attention span, working memory and problem solving), the term *augmented cognition* (AugCog) has been proposed, which can be defined as a research field that aims at supporting humans by development of computational systems that 'extend' their cognition [2].

As examples of AugCog, intelligent personal assistants exist that support humans during the execution of tasks (see e.g. [3], [4]). Such personal assistants usually include models that represent the state of the human and his or her tasks at particular time points, which can be utilized to determine when intervention is needed. An example of such a model addresses the cognitive load of the human (see e.g. [5]). The considered aspect of human behaviour and of the execution of tasks is unique. The existing models proposed for personal assistants focus on a certain domain and hence

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 3–12, 2009.

are not generic. This paper presents a generic design for a Personal Assistant agent model. The Personal Assistant can use specific dynamical models to monitor and analyse the current processes of the human. Specific sensors measure the human's psychophysiological state (e.g., heart rate) and the state of the environment (e.g., noise) to detect a possible problem and to test hypotheses. If needed, intervention actions are selected for the specific state, domain and task.

The paper is organized as follows. The generic model for a Personal Assistant agent which performs monitoring and guidance is described in Section 2. A scenario realised in a prototype implementation is described in Section 3. The multi-agent context for the Personal Assistant agent is described in Section 4. Finally, Section 5 concludes the paper.

2 The Generic Personal Assistant Agent Model

The *personal assistant agent* (PA) supports a human during the execution of a task. A personal assistant's main function is *monitoring and guidance* of the human to whom it is related. Personal assistants also interact with the physical world by performing observations (e.g., of the human's actions and their effects).The agent model for PA was designed based on the component-based Generic Agent Model (GAM) presented in [6]. Within the Generic Agent Model the component *World Interaction Management* takes care of interaction with the world, the component *Agent Interaction Management* takes care of communication with other agents. Moreover, the component *Maintenance of World Information* maintains information about the world, and the component *Maintenance of Agent Information* maintains information about other agents. The component *Own Process Control* initiates and coordinates the internal agent processes. In the component *Agent Specific Task*, domain-specific tasks were modelled, in particular monitoring and guidance. At the highest abstraction level the component consists of 5 subcomponents: *Coordination*, *Monitoring*, *Analysis*, *Plan Determination*, and *Plan Execution Preparation.*

2.1 Coordination

The initial inputs for the process are the goals provided from PA's *Own Process Control* component, which are refined within the *Coordination* component into more specific criteria that should hold for the human's functioning (e.g., 80% of certain objects on a radar screen should be identified within 30 seconds). Note that goal refinement may also occur after the initialization phase based on the results of particular observations. For example, based on the acceptance observation of a task by the human, the criteria for particular task execution states may be generated from task-related goals. More specifically, for the Personal Assistant agent a set of prioritized general goals is defined, which it strives to achieve. Some of these goals are related to the quality of the task execution, others concern the human's well-being (see Table 1). Goals of two types are distinguished:

(1) achievement goals (e.g., goals 1-3 in Table 3) that express that some state is required to be achieved at (or until) some time point, specified by

has_goal(agent, achieve(state, time))

(2) maintenance goals (e.g., goals 4-7 in Table 3) that express that some state is required to be maintained during a time interval specified by

has_goal(agent, maintain(state, begin_time, end_time))

A role description may contain role-specific goals that are added to general goals.

Although refinement may be defined for some general goals of the personal assistant agent, most of them remain rather abstract. Using the information about the human and the assigned tasks, some goals of the personal assistant agent may be refined and instantiated into more specific, operational goals. This is done by the Own Process Control component of the personal assistant agent. For example, one of the subgoals of goal 7 (*'It is required to maintain a satisfactory health condition'*) expresses '*It is required to maintain the human's heart rate within the acceptable range'*. Based on the available information about the physical characteristics of the human (e.g., the acceptable heart rate range is 80-100 beats per minute), this goal may be instantiated as '*It is required to maintain the human's heart rate 80-100 beats per minute'*. Also the task-related generic goals can be refined into more specific goals related to the particular tasks from the provided package (e.g., *'It is required to achieve the timely execution of the task repair sensor TX324'*). New goals resulting from refinement and instantiation are provided by the Own Process Control component to the Agent Specific Task component of the Personal Assistant agent, which is responsible for checking if the generated goals are satisfied. The criteria are fed to the *Monitoring* component, which is discussed below.

Table 1. General goals defined for the Personal Assistant agent

#	Goal
1	It is required to achieve the timely task execution
2	It is required to achieve a high degree of effectiveness and efficiency of the task execution
3	It is required to achieve a high degree of safety of the task execution
4	It is required to maintain the compliance to a workflow for an assigned task
5	It is required to maintain an acceptable level of experienced pressure during the task execution
6	It is required to maintain the human's health condition appropriate for the task execution
7	It is required to maintain a satisfactory health condition of the human

2.2 Monitoring

Within the *Monitoring* component, it is determined what kinds of observation foci are needed to be able to verify whether the criteria hold. In the object identification example, this could be "identification" (i.e. the event that the human identified an object).

The identified observation foci are translated into a number of concrete sensors being activated. As a form of refinement it is determined how specific information of a desired type can be obtained. For this a hierarchy of information types and types of sensors is used, as is information about the availability of sensors. For example, if the observation focus "identification" is established, the monitoring component could refine this into two more specific observation foci "start identification" and "stop identification". For the first observation an eye tracker could be turned on, while the second could be observed by looking at the events generated by a specific software component. Finally, *Monitoring* combines the detailed observations and reports the higher-level observation to *Analysis*.

2.3 Analysis

If the *Analysis* component infers (based on a conflict between the criteria and the observations) that there is a problem, it aims to find a cause of the problem. Based on an appropriate dynamic model, hypotheses about the causes are generated using forward and backward reasoning methods (cf. [7]). First, temporal backward reasoning rules are used to derive a possible hypothesis regarding the cause of the problem:

```
if      problem(at(S:STATE, I1:integers), pos)
then    derivable_backward_state(at(S:STATE, I1:integers));
if      leads_to_after(M:MODEL, S1:STATE, S2:STATE, I2:integers,pos)
  and   derivable_backward_state(at(S2:STATE, I1:integers))  and I3:integers = I1:integers - I2:integers
then    derivable_backward_state(at(S1:STATE, I3:integers));
if      intermediate_state(S:STATE)   and derivable_backward_state(at(S:STATE, I:integers))
then    possible_hypothesis(at(S:STATE, I:integers))
```

Hereby, the first rule indicates that in case a problem is detected (a state S holding at a particular time point I1), then this is a derivable backward state. The second rule states that if a causal rule specifies that from state S1 state S2 can be derived after duration I2 with a specific model (represented via the leads_to_after predicate), and the state S2 has been marked as a derivable backward state (at I1), then S1 is also a derivable backward state, which holds at I1 – I2. Finally, if something is a derivable backward state, and it is an internal state (which are the ones used as causes of problems), then this state is a possible hypothesis. Using such abductive reasoning of course does not guarantee that such hypotheses are correct (e.g. it might also be possible to derive J from another state). Therefore, the analysis component assumes one hypothesis (based upon certain heuristic knowledge, see e.g. [7]) and starts to reason forwards to derive the consequences of the hypothesis (i.e. the expected observations):

```
if      possible_hypothesis(at(S:STATE, I:integers))
then    derivable_forward_state_from(at(S:STATE, I:integers), at(S:STATE, I:integers));
if      leads_to_after(M:MODEL, S1:STATE, S2:STATE, I1:integers, pos)
  and   derivable_forward_state_from(at(S1:STATE, I2:integers),at(S3:STATE, I3:integers))
  and   I4:integers = I2:integers + I1:integers
then derivable_forward_state_from(at(S2:STATE, I4:integers), at(S3:STATE, I3:integers));
if      observable_state(S1:STATE)
  and   derivable_forward_state_from(at(S1:STATE, I1:integers), at(S2:STATE, I2:integers))
then    predicted_for(at(S1:STATE, I1:integers), at(S2:STATE, I2:integers));
```

The predictions are verified by a request from the *Monitoring* component to perform these observations. For example, if a hypothesis based on a cognitive model is that the undesired function is caused by an experienced pressure that is too high, then the observation focus will be set on the heart rate. The monitoring component selects the sensors to measure this. After these observation results come in, the selected hypothesis can be rejected in case the observations do not match the predicted observations. An example rule thereof is specified below:

```
if      observation_result(at(S1:STATE, I1:integers), neg)
  and   selected_hypothesis(at(S2:STATE, I2:integers))
  and   predicted_for(at(S1:STATE, I1:integers), at(S2:STATE, I2:integers))
then    to_be_rejected(S2:STATE);
```

Eventually, this leads to the identification of one or more specific causes of the problems, which are communicated to *Plan Determination*.

2.4 Plan Determination

Within *Plan Determination,* based on the identified causes of undesired functioning, plans are determined to remedy these causes. This makes use of causal relations between aspects in a dynamic model that can be affected and the (internal) states identified as causes of the undesired functioning. Hereby, backward reasoning methods (as explained for the *Analysis* component) are used. These use the specific cause of the problem as input, and derive what actions would remedy this cause. To decide which actions are best, the *Plan Determination* component also uses knowledge about the compatibility of solutions, their effectiveness and their side effects. See [7] for more a detailed overview of possible selection strategies. In the example, this component could conclude that the "noise level" should be reduced to lower the experienced pressure. The analysis component monitors the effectiveness of this measure. If it does not solve the problem, or causes undesired side effects, this will be considered as a new problem, which will be handled through the same process.

2.5 Plan Execution Preparation

Finally, within *Plan Execution Preparation* the plan is refined by relating it more specifically to certain actions that have to be executed at certain time points. For example, reducing the noise level could be achieved by reducing the power of an engine, or closing a door.

3 An Example Scenario

A prototype of the system has been implemented in the modelling and prototyping environment for the component-based agent design method DESIRE [8]. This prototype has been used to evaluate the model for a specific scenario as specified by domain experts of the Royal Netherlands Navy. The scenario concerns the mechanic Dave, who works on a ship of the Navy:

> *Dave just started his shift when he got an alarm that he had to do a regular check in the machine room; he accepted the alarm and walked towards the room. There he heard a strange sound and went to sit down to find the solution. However, he could not immediately identify the problem. At the same time, Dave received a critical alarm on his PDA: the close-in weapon system (CIWS) of the ship was broken. He immediately accepted the alarm, however continued to work on the engine problem, resulting in the more critical task to fix the close-in weapon system not being performed according to schedule.*

To apply the approach presented in this paper for this scenario, a number of models have been specified. First of all, the workflow models for the two tasks from the mechanic's task package have been specified. For the sake of brevity, these models are not shown, but specified in [9]. Furthermore, a cognitive model concerning the experienced pressure is specified, which is shown in Figure 1. Hereby, the nodes indicate states and the arrows represent causal relationships between these states.

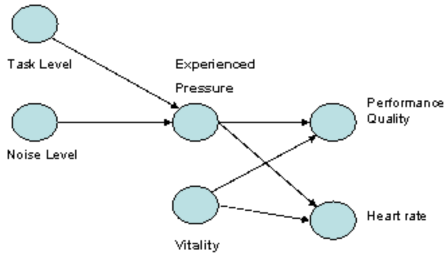


Fig. 1. Simplified cognitive model for experienced task pressure

In the agent model, relations between the states have been represented using the leads_to_after predicate, specified by means of four parameters: the model name, a condition state, a consequence state, and a delay between the two. For instance, the relation

leads_to_after(cogn1, and(normal_exp_pressure, normal_vitality), high_perf_quality, 1)

indicates that a normal experienced pressure combined with normal vitality leads to a high performance quality of the task in one step.

The presented scenario has been simulated within the prototype of the proposed architecture. Below, a brief overview of the steps the system takes is presented. When the system is started, the mechanic's task package that comprises two task types maintain_engine and solve_ciws_problem is provided to *Own Process Control* of *PA*. The mechanic is characterized by the default profile with standard characteristics (e.g., the heart rate range is 60-100 beats per minute). Furthermore, a set of generic goals provided to *Own Process Control* is defined to achieve timely task execution for each task, and to maintain a good health for the human it supports. The goal related to the mechanic's health is further refined stating that the experienced pressure and the vitality should remain normal:

own_characteristic(has_goal(PA, achieve(ontime_task_execution, -1))
own_characteristic(has_goal(PA, maintain(good_health_condition, 0, -1)))
own_characteristic(has_goal(PA, maintain(normal_exp_pressure, 0, -1)))
own_characteristic(has_goal(PA , maintain(normal_vitality, 0, -1)))

Here, '-1' indicates infinite time. Based on the goals related to the mechanic's health condition, the query for a cognitive model with the value normal_exp_pressure of the parameter states is generated and communicated by *Own Process Control* to *MMA*. As a result of this query, the model annotated by the corresponding parameters is indeed retrieved from *MMA*, and stored within the component *MAI* within *PA*:

maintenance of agent information (PA)
input: belief(leads_to_after(cogn1, and(normal_exp_pressure, normal_vitality), high_perf_quality, 1), pos) etc.
output: **see input**

The workflow models for the assigned tasks are extracted from *MMA* in a similar manner.

Eventually, the models and the goals are also received by the *Coordination* component in *Agent Specific Task*. Based on this input *Coordination* generates specific criteria. In particular, based on the goals to maintain normal_exp_pressure and normal_vitality, the criteria to maintain the medium heart rate and the high performance quality are generated using the cognitive model. The generated criteria are provided to the *Monitoring* component, which sets the observation foci corresponding for these criteria.

After this has all been done, a new assignment of a task is received from the *World* component, namely that a task of type maintain_engine has been assigned to the mechanic:

physical world
input: -
output: observation_result(at(assigned_task_at(maintain_engine, 3), 3), pos))

Based on this information *Coordination* generates new criteria using the workflow model corresponding to the task. Most of these criteria establish the time points at which the execution states from the workflow should hold, for example:

achieve(walk_to_engine, 4)

These criteria are again sent to the *Monitoring* component within *Agent Specific Task*. Therefore, the component sets the observation foci to the states within the workflow. If no goal violation is detected, no actions are undertaken by the agent. After a while however, a new task is assigned, namely the task to fix the close-in weapon system (of type solve_ciws_problem), which is outputted by the world:

observation_result(at(assigned_task_at(solve_ciws_problem, 23), 23), pos))

Again, the appropriate criteria are derived based on the corresponding workflow model. The *Monitoring* component continuously observes whether the criteria are being violated, and at time point 66 (when the mechanic should walk to the close-in weapon system) it observes that this is not the case. Therefore, a criterion violation is derived by the *Monitoring* component.

monitoring (AST - PA)
input: observation_result(at(walk_to_ciws, 66), neg); etc.
output: criterion_violation(walk_to_ciws) etc.

This criterion violation is received by the component *Analysis*, which is triggered to start analysing why the mechanic did not perform the task in a timely fashion. This analysis is performed using the cognitive model. The first hypothesis which is generated is that the cause is that the experienced pressure is normal, but the vitality abnormal. The *Analysis* component derives that a low heart rate must be observed to confirm this hypothesis (an observation that is not available yet):

analysis (AST - PA)
input: observation_result(at(walk_to_ciws, 66), neg);
criterion_violation(walk_to_ciws)
output: selected_hypothesis(at(and(normal_exp_pressure, abnormal_vitality), 65);
to_be_observed(low_heart_rate))

Since the heart rate is not observed to be low, but high, the *Analysis* component selects another hypothesis that is confirmed by the observation results that are now present (after the heart rate has been received). The resulting hypothesis is abnormal experienced pressure, and normal vitality. This hypothesis is passed on to the *Plan Determination* component within *Agent Specific Task* of the *PA* agent. *Agent Specific Task* derives that the task level should be adjusted:

plan determination (AST - PA)
input: selected_hypothesis(at(and(abnormal_exp_pressure, normal_vitality), 65)
output: to_be_adjusted(abnormal_task_level)

To achieve this adjustment, the mechanic is informed that the maintenance task is not so important, and that the mechanic should focus on the close-in weapon system task. This eventually results in a normal task level of the mechanic.

4 The Multi-agent Context for the Personal Assistant Agent

The Personal Assistant agent PA functions within the context of a multi-agent system consisting of different types of agents. In addition to the Personal Assistant itself the following agents are involved; models for all of them were designed based on the component-based Generic Agent Model (GAM) presented in [6]. The *Model Maintenance Agent* (MMA) contains a library of four types of models: monitoring and guidance models, cognitive models, workflow models and dialogue models. Models can be provided to PA upon request; to facilitate this process, each model is annotated with specific parameters. The *State Maintenance Agent* (SMA) maintains characteristics, states and histories of other agents, of the physical world and of the workflows. Information can be requested by the PA's, using a specific element (i.e. agent, physical world, a workflow), an aspect (i.e. state, history) and a time interval for which information should be provided. In addition, the *Mental Operations Agent* (MOA) represents the mental part of the human. MOA is connected to the human's physical body, which can act in the physical worlds. The *Task Execution Support Agent* (TESA) is used by the human as an (active) tool during the execution of a task.

For each human that needs to be supported during the task execution a Personal Assistant agent is created. Initially, the Personal Assistant agent contains generic components only. The configuration of it is performed based on the role that needs to be supported by the agent, on the characteristics of a human who is assigned to this role, and on the goals defined for the Personal Assistant agent.

The configuration of the self-maintaining personal assistant agent begins with the identification of the suitable monitoring and guidance task model(s) that need(s) to be requested from the model maintenance agent. To this end, the model parameters are identified by the Own Process Control component based on the goals of the personal assistant agent. For example, to establish if the human complies with a workflow model, diagnosis of the human's state may need to be performed. Thus, a query to the model maintenance agent is given which includes the parameter type of analysis with value diagnosis. When a query is specified, the function model_query(query_id, param, list_of_values) is used, where the first argument indicates a query identifier, the second argument indicates a parameter and the third argument indicates a list of parameter values.

The choice of cognitive models is guided by the goals that concern internal states of the human. From the goals in Table 1 and their refinements and instantiations, a number of internal states can be identified, among which experienced pressure and heart rate. For such states and for each task the appropriate cognitive, workflow and dialogue models are extracted from the model maintenance agent. By matching queries received from the personal assistant agent with the annotations of the maintained models, the model maintenance agent identifies the most suitable model(s), which is (are) communicated to the requestor. The provided models are stored in the Maintenance of Agent Information component of the personal assistant.

More details about the multi-agent context of the personal assistant agent can be found in [10].

5 Conclusions

In every organisation a set of critical tasks exists that greatly influence the satisfaction of important organisational goals. Thus, it is required to ensure effective and efficient execution of such tasks. To this end, automated personalized assistance for the task performers may be used. In this paper, a generic agent model for personal support during task execution has been proposed. This agent model allows the use of dynamical models and information about the assigned goals and tasks. The personal assistant agent performs monitoring and analysis of the behaviour of the supported human in his/her environment. In case a known problem is detected, the agent tries to identify and execute an appropriate repair action. The fact that the architecture is generic differentiates the approach from other personal assistants such as presented in [5; 6]. Besides being generic, the proposed personal assistant agent has an advantage of being relatively lightweight, as it only maintains and processes those models that are actually needed for the performance of the tasks. It can therefore run upon for instance a PDA or cell phone. To provide the required functionality for personal assistant agents, the multi-agent context in which it functions includes model maintenance and state maintenance agents.

When performing a task, especially in highly demanding circumstances, human performance can be degraded due to increased cognitive workload. A possible negative effect of high cognitive workload is that it leads to a reduction in attention and situation awareness [11]. Situation awareness refers to the picture that people have of the environment (e.g., [12]). In case of low situation awareness this picture is wrong, which will often lead to wrong decision making (e.g., [13]). In the literature, it is known that automated systems can also impose a negative effect on cognitive workload or situation awareness [14]. Therefore, systems have been designed that are adaptive, e.g. in only providing aiding when it is necessary [5]. For this, a human's cognitive state should be assessed online; since this is difficult, often adaptive systems like this are based on psychophysiological measurements, like brain activity and eye movements (e.g. [15], [5]). The personal assistant model described in this paper makes use of such measurements, but in addition uses models of cognitive states and dynamics, and the current workflow to be able to assess the online state of the human. This allows for an optimal support of the human.

References

1. Posner, M.I., Boies, S.J.: Components of attention. Psychological Bulletin 78, 391–408 (1971)
2. Schmorrow, D.D., Reeves, L.M.: 21st century human-system computing: augmented cognition for improved human performance. Aviat Space Environ. Med. 78(Suppl. 5), B7–B11 (2007)
3. Myers, K., Berry, P., Blythe, J., Conley, K., Gervasio, M., McGuinness, D.L., Morley, D., Pfeffer, A., Pollack, M., Tambe, M.: An Intelligent Personal Assistant for Task and Time Management. AI Magazine, 47–61 (Summer 2007)

4. Modi, P.J., Veloso, M., Smith, S.F., Oh, J.: CMRadar: A Personal Assistant Agent for Calendar Management. In: Bresciani, P., Giorgini, P., Henderson-Sellers, B., Low, G., Winikoff, M. (eds.) AOIS 2004. LNCS, vol. 3508, pp. 169–181. Springer, Heidelberg (2005)
5. Wilson, G.F., Russell, C.A.: Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding. Human Factors 49(6), 1005–1018 (2007)
6. Brazier, F.M.T., Jonker, C.M., Treur, J.: Compositional Design and Reuse of a Generic Agent Model. Applied AI Journal 14, 491–538 (2000)
7. Duell, R., Hoogendoorn, M., Klein, M.C.A., Treur, J.: An Ambient Intelligent Agent Model using Controlled Model-Based Reasoning to Determine Causes and Remedies for Monitored Problems. In: Proceedings of the Second International Workshop on Human Aspects in Ambient Intelligence, HAI 2008. IEEE Computer Society Press, Los Alamitos (2008)
8. Brazier, F.M.T., Jonker, C.M., Treur, J.: Principles of Component-Based Design of Intelligent Agents. Data and Knowledge Engineering 41, 1–28 (2002)
9. `http://www.cs.vu.nl/~wai/PersonalAssistant/Models.pdf`
10. Bosse, T., Duell, R., Hoogendoorn, M., Klein, M.C.A., Lambalgen, R., van Mee, A., van der Oorburg, R., Sharpanskykh, A., Treur, J., de Vos, M.: A Multi-Agent System Architecture for Personal Support During Demanding Tasks. In: Proc. of the 22nd Int. Conf. on Industrial, Engineering & Other Applications of Applied Intelligent Systems, IEA/AIE 2009. Studies in Computational Intelligence. Springer, Heidelberg (to appear, 2009)
11. Wickens, C.D.: Situation awareness and workload in aviation. Current Directions in Psych. Science 11, 128–133 (2002)
12. Endsley, M.R.: Theoretical underpinnings of situation awareness. In: Endsley, M.R., Garland, D.J. (eds.) Situation awareness analysis and measurement, pp. 1–21. Erlbaum, Mahwah (2000)
13. Endsley, M.R.: The role of situation awareness in naturalistic decision making. In: Zsambok, C., Klein, G. (eds.) Naturalistic decision making, pp. 269–284. Erlbaum, Mahwah (1997)
14. Parasuraman, R., Riley, V.: Humans and automation: use, misuse, disuse, abuse. Human Factors 39(2), 230–253 (1997)
15. Prinzel, L.J., Freeman, F.G., Scerbo, M.W., Mikulka, P.J., Pope, A.T.: A closed-loop system for examining psychophysiological measures for adaptive task allocation. Int. Journal of Aviation Psychology 10(4), 393–410 (2000)

Adaptive Interfaces in Driving

Rino F.T. Brouwer[1], Marieka Hoedemaeker[2], and Mark A. Neerincx[1,2]

[1] TNO Human Factors, Kampweg 5, 3769 ZG Soesterberg, The Netherlands
[2] Delft University of Technology, Mekelweg 4, 2628 GA Delft, The Netherlands
rino.brouwer@tno.nl, marieka.hoedemaeker@tno.nl, mark.neerincx@tno.nl

Abstract. The automotive domain is an excellent domain for investigating augmented cognition methods, and one of the domains that can provide the applications. We developed, applied and tested indirect (or derived) measures to estimate driver state risks, validated by direct state-sensing methods, with major European vehicle manufacturers, suppliers and research institutes in the project AIDE (Adaptive Integrated Driver-vehicle InterfacE). The project developed an interface with the driver that integrates different advanced driver assistant systems and in-vehicle information systems and adapted the interface to different driver or traffic conditions. This paper presents an overview of the AIDE project and will then focus on the adaptation aspect of AIDE. Information presented to the driver could be adapted on basis of environmental conditions (weather and traffic), and on basis of assessed workload, distraction, and physical condition of the driver. The adaptation of how information is presented to the driver or the timing of when information is presented to the driver is of importance. Adapting information, however, also results in systems that are less transparent to the driver.

Keywords: In-car services, workload, adaptive user interface, central management.

1 Introduction

A major research effort on augmented cognition takes place in the defense domain, aiming at systems that support or extend the limited human information processes for operations in high-demand situations [1]. To augment cognition in dynamic conditions, the momentary human state is often sensed via (psycho)physiological measurements, such as EEG and heart rate [2]. New non-obtrusive methods can be used, such as camera sensors and microphones to assess emotion out of, respectively, facial expressions and voice [3]. In general, we propose to use a mixture of methods, including measures of human, task and context [4]

In our view, the automotive domain is an excellent domain for investigating augmented cognition methods, and one of the domains that can provide the applications. First, the human is in a constrained (relatively fixed, “indoor”) position, sitting in an environment that can be relatively easily enriched with driver-state sensing technology. Second, the driver’s tasks is rather well-defined, and can be tracked well, and

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 13–19, 2009.

context factors can be easily assessed via both current sensor technology (e.g., slippery road) and data acquisition via wireless networks (e.g., traffic density and weather). These domain and task characteristics allow for high-levels of automation to support safety and comfort, but the human task performance will remain a crucial factor of the overall driver-car performance. Third, there seems to be a real need for AugCog technology. Drivers can access more and more services in the car, for example for navigation, traffic information, news and communication. Furthermore, the car itself provides more and more information that should support drivers' tasks, such as speed limit warnings and parking guidance "beeps". The consequences of providing in-car traffic management information (like route information) in combination with infotainment services (like news headlines) can be negative; distraction or high workload could adversely affect the interaction between the driver and the in-car system (e.g. [5], [6]). Overload means that the driver is unable to process all relevant information necessary to perform the primary driving task. This may lead to increased error rates and delayed detection of other traffic participants and, hence, to reduced safety [7].

A recent study showed that 93% of observed crashes related to 'inattention' [8]. Within traffic research detection of 'inattention' (eyes not on the road) plays an important role. The 'eyes not on the road' can be caused by many things such as distraction, drowsiness, intoxication, workload, etc. It is not an easy task to detect 'inattention'. Clearly drowsiness can be detected through EEG signals but no driver will step into a car and puts an EEG cap on. So alternative measures needed to be developed. A lot of research effort was put into developing such measures. However still none provided a detection good enough to develop an in-vehicle system. The number of accidents is the measure for traffic safety. Although they happen on a daily basis accidents are fortunately still quite rare. So also with respect to the traffic safety alternative measures or indicators are needed. In traffic research, objective measures were developed that relate to the lateral part (e.g., how does a driver keeps its lane) and the longitudinal part of the driving task (e.g., car following). Of some of these measures it could be shown that there was a correlation between the measurement (e.g., speed) and traffic safety [9]. Other measures such as the duration until a driver crosses a line marking given the same speed and acceleration (time-to-line crossing) or the time-to-collision have also shown to be related to traffic safety. Subjective questionnaires were developed to indicate workload experienced by the driver. However under normal driving conditions it is unwise to fill out a questionnaire to assess the workload of the driver. So objective measures were used that a related to the steering behaviour of the driver (such as steering reversal rate). An extended list of measures that are commonly used in traffic research was generated by the AIDE project (e.g., [6]).

The importance of measuring the status of the driver (workload, distraction, etc) while driving lies in the possibility to warn a driver for potential hazardous situations and for adapting the interface to the driver. A driver that is distracted will need an earlier warning of a system in order to avoid a possible collision then a driver who is not distracted. However adapting the HMI to the driver requires storing some data of that driver. So adapting the HMI brings along privacy issues (e.g., who has access to the stored data). Also the introduction of driver support systems brings along other problems then just technical or HMI related. For example, an adaptive cruise control (ACC) can not only maintain a certain speed but also a certain distance to a leading vehicle. If that vehicle drives slower then the ACC vehicle then the ACC vehicle has to slow down too. However this deceleration is limited. If the leading vehicle suddenly brakes harsh then the ACC might technically able to cope but this cannot be guaranteed for all kind of situations. To avoid such legal issues on who is the blame in case of an accident when there are driver assistance systems on board, it is always stated that the driver is responsible, meaning should always stay in the loop with respect to the driving task.

To address all appplication constraints of AugCog technology, the AIDE project developed, applied and tested alternative (or derived) measures to estimate driver state risks. In this approach, the direct state-sensing methods (like eye-tracking and hear rate) are used to validate these measures.

2 The AIDE Project

Within Europe in 2007 about 43000 people died as the consequence of a traffic accident and about 1.7 million people were injured. Human error is the main contributing factor in accidents. To assist drivers in their task Advanced Driver Assistance Systems (ADAS such as forward collision warning systems, lane departure warning systems, vision enhancement systems) have been developed that offer great potential for improving road safety. These systems can warn the driver with respect to (potential) dangerous situations but can also to a certain extent take over part of the driving task. In-vehicle information systems only inform the driver and are most of the time not directly related to the driving task (e.g., mobile phone, fleet management, but also route navigation). Although these systems have benefits either with respect to driving safety or comfort there is huge risk that if the systems work in isolation the workload of the driver may increase thereby compromising traffic safety. Integration and adaptation of the systems are important tools to have the benefits of these systems without having the side effects. The AIDE project (Adaptive Integrated Driver-vehicle interfacE; IST-1-507674-IP) wanted to generate the knowledge and develop methodologies and human-machine interface technologies required for safe and efficient integration of ADAS, IVIS and nomad devices into the driving environment. The objectives of AIDE are

- to maximize the efficiency, and hence the safety benefits, of advanced driver assistance systems,
- to minimize the level of workload and distraction imposed by in-vehicle information systems and nomad devices and
- to enable the potential benefits of new in-vehicle technologies and nomad devices in terms of mobility and comfort.

To reach the objectives an integrated HMI was developed and tested in which the following components was developed

- Multimodal HMI I/O devices shared by different ADAS and IVIS (e.g. head-up displays, speech input/output, seats vibrators, haptic input devices, directional sound output)
- A centralised intelligence for resolving conflicts between systems (e.g. by means of information prioritisation and scheduling).

- Seamless integration of nomadic devices into the on-board driver-vehicle interface.
- Adaptivity of the integrated HMI to the current driver state/driving context. The adaptive interface should also be re-configurable for the different drivers' characteristics, needs and preferences. This requires techniques for real-time monitoring of the state of the driver-vehicle-interface system.

To illustrate best what AIDE aimed at is the vision that was laid down in the AIDE proposal:

> "Maria starts the car and drives through the city centre towards the motorway that leads to the small seaside town where she lives. When the car starts moving, all functions not suitable for use while driving are disabled. It is rush hour and the streets are crowded with other vehicles, pedestrians and bicyclists.
>
> By means of using information gathered from on-board sensors combined with a satellite-based positioning system, the car knows that the driving situation is demanding and adapts the driver-vehicle interface so that Maria can concentrate on the driving. Thus, the information given through the interface is reduced to a minimum and all non-critical information is put on hold until later. Moreover, irrelevant safety systems, e.g. lateral control support, are disabled.
>
> When Maria stops at a traffic light a voice message is given informing her that the road ahead is blocked and suggests an alternative route. This message was judged by to be sufficiently important to be let through despite the overall demanding driving context, but the system waited to present it until the workload was temporary reduced at the traffic light.
>
> After driving for a few minutes on the highway, Maria starts thinking about a complex lawsuit that she has been assigned the responsibility for at work. The vehicle detects the increased cognitive activity from changes in her eye-movement patterns (detected by the cameras in the dashboard). After a while, the vehicle in front of hers brakes for a traffic queue. This is detected by the collision avoidance system, which alerts Maria of the potential danger using a flashing light combined with a slight seat vibration. She gets the alert well in time to be able to avoid the danger. However, since Maria was cognitively distracted, the warning was given earlier and the intensity of the warning was stronger than would have been the case if Maria had been fully attentive."[1]

Clearly not everything can not yet be implemented but for example adjusting the HMI based on "satellite-based positioning system" can easily be achieved. Within AIDE three different prototypes were developed: One truck and two cars.

An example: Adapting a forward collision warning system

This paper focuses on the adaptivity aspect of the AIDE project and more precisely on the acceptance of an adaptive system.[2] In AIDE a large number of experiments were performed with respect to the different aspects of the AIDE system. Three closely

---

[1] Taken from the AIDE website http://www.aide-eu.org/index.html

[2] For more information on the AIDE project the interested reader is referred to the AIDE IP website (http://www.aide-eu.org/index.html) or you can contact Rino Brouwer at rino.brouwer@tno.nl

related experiments were performed by ITS Leeds (UK), VTI (Sweden) and TNO. In these experiments the effects of a Forward Collision Warning system were investigated. A Forward Collision Warning (FCW) is an on-board electronic safety device that continuously monitors traffic obstacles in front of the host vehicle and warns the driver when a risk of collision is imminent. The benefits of an FCW in reducing the number and severity of front-to-back collisions or ‘shunts’ have been reported (e.g, [10]). The effects of the system on driving behavior and on acceptance of the system were investigated in three driving simulator experiments (see Figure 1). In the experiment performed by ITS Leeds the FCW was adapted to the driver, in the experiment of VTI it adapted to the road friction, and in the experiment by TNO to distraction.

Fig. 1. The driving simulators used in the experiments. Top left, the TNO simulator; bottom left, the (old) ITS Leeds simulator; right the moving base driving simulator at VTI.

In all three experiments participants had to drive a route of 40 km in which a leading vehicle could sometimes suddenly brake in which the FCW could give a warning. In all experiment driving with an adapted FCW was compared to driving without an adapted FCW. As stated at ITS Leeds the system was adapted to individual differences. For drivers with a short reaction time the system warned later then for drivers with a longer reaction time. At VTI the FCWwas adapted whether the roads was slippery or not. In case of a slippery road the system warned earlier than on a dry road. At TNO the FCW warned earlier when the driver was distracted which was achieved by letting the driver perform a secondary task (for more detailed information on these experiments see [11]).

User acceptance was assessed by using the Van der Laan scale [12], giving a rating for satisfaction and usefulness of each FCW type. This scale consist of nine questions which reflect the underlying scale satisfaction and usefulness. (see Table 1).

Table 1. The questions in the van der Laan scale

<table>
<tr><td>Useful</td><td>|__| __| __| __| __|</td><td>Useless</td></tr>
<tr><td>Pleasant</td><td>|__| __| __| __| __|</td><td>Unpleasant</td></tr>
<tr><td>Bad</td><td>|__| __| __| __| __|</td><td>Good</td></tr>
<tr><td>Nice</td><td>|__| __| __| __| __|</td><td>Annoying</td></tr>
<tr><td>Effective</td><td>|__| __| __| __| __|</td><td>Superfluous</td></tr>
<tr><td>Irritating</td><td>|__| __| __| __| __|</td><td>Likeable</td></tr>
<tr><td>Assisting</td><td>|__| __| __| __| __|</td><td>Worthless</td></tr>
<tr><td>Undesirable</td><td>|__| __| __| __| __|</td><td>Desirable</td></tr>
<tr><td>Raising Alertness</td><td>|__| __| __| __| __|</td><td>Sleep-inducing</td></tr>
</table>

The results for the three experiments showed that only the adaptive FCW in the experiment of Leeds was rated more positively then the non-adaptive FCW. In both experiments of VTI and TNO the non-adaptive system was rated more positively. Although there are some differences between the three experiments an important difference was that in the experiment of ITS Leeds the system was adapted to individual differences while at VTI and TNO the system was adapted to circumstances (slippery roads or distraction). The adaptation of the system to a driver's preference is more likely to be noticed by the driver then a system that adapts to circumstances. Although the road may look slippery it may not be clear to the driver that the system warns earlier because of less friction. And although the driver has to perform a secondary task and is distracted (at least that is assumed) the driving task might still be manageable together with the secondary task. So it may not clear to the driver why the system warns earlier. In both the friction and the distraction experiment the driver may only perceive that a warning is given earlier but not why.

3 Conclusions

This paper presented an approach to realize "Augemented Cognition" in a car by adaptive in-car information and service presentations. According to this approach critical user states are assessed via context information, and validated in high-fidelity driver simulators. Via sensing the driver behaviour, information provision and environmental conditions, the actual critical states can be detected, and the in-car interfaces can be changed to establish adequate load levels. The most important developments in this area are the Advanced Driver Assistance Systems (ADAS) and In Vehicle Information Systems (IVIS) [11].

The AIDE project showed that information presented to the driver could be adapted on basis of environmental conditions (weather and traffic), and on basis of assessed workload, distraction, and physical condition of the driver [13]. The adaptation of how information is presented to the driver or the timing of when information is presented to the driver proved to be of importance. Adapting information, however, also proved to result in systems that are less transparent to the driver. Tests in the driver simulators showed that the rationale of adaptation, such as assumed distraction, is not always clear for the drivers, resulting in less acceptance. Actually, the drivers may have to learn that the circumtances and own state bring about a safety risk, and feedback on this aspect might help to improve the acceptance. In other words, the adaptive interface should explain its behaviour (e.g., during a training session). Furthermore, the experiments showed that personalization can be beneficial on this aspect.

References

1. Schmorrow, D.D., Nicholson, D.M., Drexler, J.M., Reeves, L.M. (eds.): Proc. 4th Intern. Augmented Cognition. Strategic Analysis, Inc., Arlington (2007)
2. Voskamp, J.: Measuring cognitive workload in non-military scenarios: criteria for sensor technologies. In: Proceedings of the 5th International Conference of the Augmented Cognition, San Diego, CA, USA, July 19-24 (2009)
3. Truong, K.P., van Leeuwen, D.A., Neerincx, M.A.: Unobtrusive Multimodal Emotion Detection in Adaptive Interfaces: Speech and Facial Expressions. In: Schmorrow, D.D., Reeves, L.M. (eds.) HCII 2007 and FAC 2007. LNCS (LNAI), vol. 4565, pp. 354–363. Springer, Heidelberg (2007)
4. Grootjen, M., Neerincx, M.A., Stolk, K.D., Weert, J.C.M.: Design and user evaluation of an interface prototype that adapts to the operator's cognitive task load. In: Schmorrow, D.D., Nicholson, D.M., Drexler, J.M., Reeves, L.M. (eds.) Proceedings of the 4th International Conference of the Augmented Cognition, pp. 97–106. Strategic Analysis, Inc., Arlington (2007)
5. McKnight, J., McKnight, S.: The effect of cellular phone use upon driver attention. Accidents Analysis and Prevention 25, 259–265 (1991)
6. Hoedemaeker, M., Neerincx, M.A.: Attuning In-Car User Interfaces to the Momentary Cognitive Load. In: Schmorrow, D.D., Reeves, L.M. (eds.) HCII 2007 and FAC 2007. LNCS, vol. 4565, pp. 286–293. Springer, Heidelberg (2007)
7. Verwey, W.B.: Evaluating safety effects of traffic congestion information systems. In: Hancock, P.A., Desmond, P. (eds.) Stress, workload and fatigue, pp. 409–425. Erlbaum, Mahwah (2001)
8. Klauer, S.G.: The impact of driver inattention on near-crash/crash risk: an analysis using the 100-car naturalistic driving study data. National Highway Traffic Safety Administration, Washington (2006)
9. Nilsson, G.: Traffic safety dimensions and the power model to describe the effect of speed on safety. Doctoral Thesis Lund Institute of Technology. Sweden (2004)
10. Lee, J.D., McGehee, D.V., Brown, T.L., Reyes, M.L.: Human Factors 44(2), 314–334 (2002)
11. Lenior, D., Janssen, W., Neerincx, M.A., Schreibers, K.: Human Factors Engineering for Smart Transport: Decision Support for Car Driver and Train Traffic Controller. Applied Ergonomics. Special edition Meeting Diversity in Ergonomics 37(4), 479–490 (2006)
12. Van Der Laan, J.D., Heino, A., De Waard, D.: A simple procedure for the assessment of acceptance of advanced transport telematics. Transportation Research Part C: Emerging Technologies 5(1), 1–10(13) (1997)
13. Verwey, W.B.: On-line driver workload estimation. Effects of road situation and age on secondary task measures. Ergonomics 43, 187–209(8) (2000)

Using Context to Identify Difficult Driving Situations in Unstructured Environments

Kevin R. Dixon[1], Justin D. Basilico[1], Chris Forsythe[1], and Wilhelm E. Kincses[2]

[1] Sandia National Laboratories, Albuquerque, NM, 87185 USA
{krdixon,jdbasil,jcforsy}@sandia.gov
[2] Daimler AG Group Research, 71059 Sildelfingen, Germany
wilhelm.kincses@daimler.com

Abstract. We present a context-based machine-learning approach for identifying difficult driving situations using sensor data that is readily available in commercial vehicles. The goal of this system is improve vehicle safety by alerting drivers to potentially dangerous situations. The context-based approach is a two-step learning process by first performing unsupervised learning to discover meaningful regularities, or "contexts," in the vehicle data and then performing supervised learning, mapping the current context to a measure of driving difficulty. To validate the benefit of this approach, we collected driving data from a set of experiments involving both on-road and off-road driving tasks in unstructured environments. We demonstrate that context recognition greatly improves the performance of identifying difficult driving situations and show that the driving-difficulty system achieves a human level of performance on cross-validation data.

1 Introduction

Cars are an essential means of transportation for much of the world. However, the widespread use of automobiles exacts a large toll in the form of property damage, injury, and death. The United States National Highway Traffic Safety Administration reports that "In 2005, there were an estimated 6,159,000 police-reported traffic crashes, in which 43,443 people were killed and 2,699,000 people were injured;" it is the leading cause of death of people aged 3 through 33 [1]. Naturalistic driving studies have shown that having a passenger in the vehicle reduces the odds-ratio of having a crash by 50% [2]. The goal of this research is not to automate driving, but to identify and mitigate potentially dangerous situations for the driver, similar to a "backseat driver," improving safety. To this end, we have conducted a series of experiments in both on-road and off-road driving in unstructured environments. In these experiments, we have shown that our system identifies difficult driving situations with performance similar to that of a human backseat driver, and see significant improvements in the performance of drivers during the experimental conditions. Our driving-difficulty classifier system operates in real time in unstructured environments without human intervention, using sensors that are readily available on commercial vehicles without additional instrumentation.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 20–29, 2009.

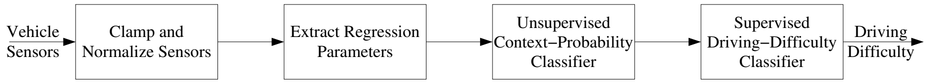


Fig. 1. Data-flow diagram in the context-based difficulty classifier system

We create the driving-difficulty detector using a two-step semi-supervised machine-learning approach [3]. The first step takes unlabeled data from the vehicle's data bus and automatically extracts the context by automatically identifying statistical regularities in the vehicle data. Our hypothesis is that the driver performing the underlying physical task - driving in the given conditions - induces observable regularities in the vehicle data and identifying these regularities, or "contexts," is crucial in achieving a human-level of performance. For example, entering a high-speed roadway tends to result in a driver pressing down the accelerator pedal, entering a period of relatively high lateral acceleration, turning on a lane-change signal, and achieving a fast speed. In this example, the underlying physical task induces regularities in how the driver interacts with the vehicle. We are interested in automatically extracting contexts to determine when the driver is entering a potentially difficult situation. With the contexts identified, the system then maps these contexts onto a difficulty score using a supervised-learning machine-learning algorithm (Fig. 1). To validate the system, we compare the performance of an actual human backseat driver with our automated system, both with and without context recognition, in identifying potentially dangerous driving conditions.

2 Related Work

For over twenty years, there has been interest in developing autonomous driving systems, with an early example being the NAVLAB project [4] and research is ongoing [5]. Autonomous driving systems have recently gained widespread attention in the research community and mainstream media, due in large part to the DARPA Grand Challenge [6] and the follow-on DARPA Urban Grand Challenge. While computer systems and robots may one day replace humans as the main users of the world's highways, it is likely that humans will continue to be the primary drivers of motor vehicles for the near future. This will continue the trend of over 40,000 fatalities per year in the United States alone, coupled with incalculable related damages [1]. The 100-car naturalistic driving study [2] recorded almost 10,000 crashes, near crashes, and "crash-relevant conflicts" over the course of about one year. This averages to about seven incidents per subject per month. One bright spot is that the same study showed that having a passenger in the vehicle reduces the odds-ratio of having a crash by 50% [2]. In some sense, the goal of this research is to have the same crash-reducing effect that passengers had in the naturalistic driving study. There has been substantial research into driver-assistance systems. Many systems focus on placing additional sensors on the vehicle, particularly visible-light cameras [7, 8], to identify previously undetectable situations. Other groups have focused on developing models of human drivers to focus attention [9]. While these are very promising avenues to pursue, we feel that we can offer powerful driver-assistance tools by intelligently analyzing readily available sensors on commercial vehicles to determine how the current situation can impact driver performance. Unsupervised learning has been used as a basis for context recognition for mobile devices [10] and for improving image classification [11]

The work presented in this paper extends the previous work in driving-difficulty systems of [12], which trained a classification system to identify potentially dangerous driving conditions using predefined situations. This system identified eight high-level situations with high accuracy: 1) Approaching or Waiting at Intersection, 2) Leaving Intersection, 3) Entering On-ramp or High-Speed Roadway, 4) Being Overtaken, 5) High Acceleration or Dynamic State of Vehicle, 6) Approaching Slow-Moving Vehicle, 7) Preparing to Change Lanes, and 8) Changing Lanes. However, this system was based purely on supervised-learning classifications on predefined categories. The primary limitation is that predefined categories are inherently limited by the cleverness of the developers to identify all relevant situations, while ignoring irrelevant ones. This also means that the system must have numerous examples of each situation against which to train the classifier. Out of the 24 hours of data collected, the rarest situation, "Entering On-ramp or High-Speed Roadway," was present for less than 1% of the data and it is very challenging for any machine-learning classifier to identify rare events [13]. Building on this previous work, our system uses a two-stage approach to identifying potentially dangerous driving conditions.

3 Algorithms

The central component of our approach is the automated unsupervised learning of context. Because we typically have a much larger amount of unlabeled data than labeled data, we take a semi-supervised approach to learning. The creation of contexts using unsupervised clustering algorithms makes use of all data recorded from an experimental vehicle. The supervised learning of driving difficulty makes use of the smaller amount of labeled data. This allows the driving-difficulty classifier to make productive use of all the unlabeled and labeled data.

3.1 Data Representation

The input to the system is a discrete-time temporal signal, which is extracted from sensors aboard an experimental vehicle from its standard Controller Area Network (CAN) bus (Section 4.1). Because we are interested in the change of the sensor values over time, we extract the rate-of-change and current-value information from each signal over a fixed time window. This feature-extraction process converts temporal signals into a vector-based representation. In terms of the features to use in the driving-context recognition, we feel that:

1. The magnitude of a signal is important. For example, knowing the speed of the vehicle or brake-pedal force can help to disambiguate similar contexts.
2. The general trend of a signal is also important. For example, knowing how sensors are changing can differentiate otherwise identical contexts.

With this in mind, at each time step for each input sensor, we construct a window over some predefined length into the past (typically 5 seconds) and compute the first-order linear-regression slope-intercept coefficients $\{m,b\}$ for that time window.

Converting a windowed temporal signal into a vector using the linear-regression coefficients creates two coordinates; the regression slope (m) and the regression intercept (b). Consequently, if there are 5 input signals, the result will be a 10-dimensional vector. Our unsupervised-clustering algorithms search for driving contexts in this vector space.

3.2 Unsupervised Context Learning

At each time step, the input to the unsupervised-learning context classifier is the collection of vectors with the slope-intercept regression parameters for each sensor. The unsupervised context-learning algorithm is a reductionist version of the prevalent k-means clustering algorithm [3]. To determine vector similarity, we use the Mahalanobis distance and compute the sample mean and full covariance matrices belonging to each cluster. We make an assumption that each regression-coefficient vector is generated independently of all others. With this assumption, the number of data points assigned to a *particular* cluster is a binomial random variable, and we remove a cluster if its corresponding probability is too low. By evaluating the binomial cumulative distribution function, we can determine if a cluster is not significant, in a statistical sense, and should be removed. If we have k clusters and N data-points, then the expectation is that each cluster contains N/k data-points. From this perspective, we can set a removal threshold based on the fraction of data-points of the expectation. For example, a threshold of 0.5 means that we will remove any clusters containing less than $0.5N/k$ data-points. In practice on our experimental data, this reductionist clustering approach yields relatively stable numbers of clusters from random initializations ($E\{k\}$=53.5,±1.92, $p<0.05$ for a removal threshold of 0.5). We also find the reductionist clustering approach to less sensitive to the initial parameter k because if the value of k is initially set too high, the algorithm will compensate by removing spurious clusters. Thus, to set k we can initially choose a relatively high value and then let the algorithm iteratively remove clusters to find a stable value.

3.3 Supervised Learning of Driving Difficulty

Up to this point, the system has mapped temporal vehicle sensors to a k-dimensional vector of context probabilities (cf. Fig. 1). We use supervised learning to map this context-probability vector to a difficulty score. As we describe later in more detail in Section 4.1.2, we collected labels of driving difficulty for a subset of the experimental data, by either backseat observation or *post hoc* video analysis. We use these scalar 1-100 value labels as ground-truth outputs for a supervised-learning algorithm. Because the values are continuous, this difficulty classifier can be stated as a standard regression formulation. Not surprisingly, driving difficulty does not change dramatically from second to second and the ground-truth difficulty labels are highly autocorrelated (R=0.89 at 5-second lag).

4 Experimental Description

We have conducted a series of driving experiments in unstructured environments over the past several years. The first studies were a proof of concept that we could infer

Fig. 2. Frontal camera view from the Camp Pendleton experiments used for *post hoc* labeling and analysis

difficult driving situation from readily available sensors from a commercial vehicle in naturalistic on-road driving conditions [12]. The set of experiments covered by this analysis involved driving in off-road conditions, on semi-improved and unimproved paths, at the United States Marine Corps Base Camp Pendleton. These experiments tested the ability for our system to identify high-difficulty driving conditions without the presence of human-made regularities, such as traffic lights, lanes, and signage. Drivers were instructed to drive on a predefined road circuit, but we did not attempt to alter the roadway and or control external conditions in any of the experiments. As such, we have encountered snow, rain, fog, traffic jams, road construction, mechanical problems, armed guards, artillery howitzers, lost vehicles, and even flocks of sheep (complete with over-protective herding dogs). Through the evolution of these experiments and the knowledge gained, we have learned that identifying driving *context* is crucial in achieving human-level accuracy with a driving-classification system. By context, we mean those regularities that are caused by the human operator (the driver) making the vehicle behave in a constrained manner.

4.1 Data Collection

Before each experiment, the subjects familiarize themselves with the test vehicle and drive on a sample course. Additionally, before the main experiment, we conduct a calibration study where we collect data from a small number of subjects with which we train our difficulty-classification system. The purpose of the calibration study was to duplicate the experimental conditions and gain insight into the phenomena that would be helpful in identifying high-difficulty situations. The use of calibration data also meant that a general driving-difficulty model is used, rather than a unique model for each driver. After the calibration study is complete, the main pool of subjects performs the driving study, as in [14].

Vehicle Data. To obtain information about the state of the vehicle and how the driver is interacting with it, we interfaced through the Controller Area Network (CAN) bus

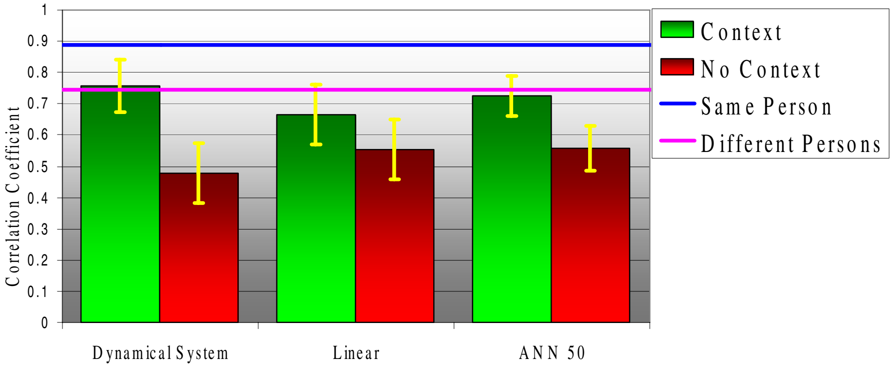


Fig. 3. Performance of the difficulty classifier with and without context recognition in terms of normalized correlation coefficient (ρ). The error bars are the cross-subject 95% confidence intervals. Same persons have an average of ρ=0.95 for subsequent scoring of the same round. Different persons difficulty scores have an average of ρ=0.74 agreement.

of the vehicle to record sensors that are readily available on many commercial vehicles. In all experiments, we sampled data from the CAN bus at 4Hz. In our experience, sampling rates above 4Hz did not improve performance, and sampling slower than 2Hz could result in missed events. Our difficulty-classification system incorporated two types of sensors: sensors that directly measure how the driver is interacting with the vehicle and sensors that measure secondary interactions or vehicle state. From the control-surface state, we made use of steering-wheel position, force applied to the brake pedal, and accelerator-pedal deflection. From the physical state of the vehicle, we made use of wheel speeds, adaptive cruise-control radar, and current gear number. There are many driver-assistance systems that require special-purpose instrumentation [8] and these provide valuable insight into the cost-benefit analysis of additional instrumentation to vehicles. However, our driving-difficulty classifier does not require any experiment-specific instrumentation of the driver or vehicle, meaning that this system is deployable on currently available commercial vehicles.

Difficulty Labels. To generate the ground-truth labels, the difficulty of the current driving situation were scored on 100-point scale (1 to 100) entered by a human labeler with an external dial or a software slider bar. A value of 1 means that the driving is very easy, while a value of 100 means that there is imminent danger. Furthermore, the labelers were instructed that a score of 50 or above indicated a judgment that it would be a bad time to burden the driver with additional tasks, such as a mobile-phone call. Allowing the labelers to input a continuous value on a 100-point scale, instead of a binary difficulty decision, makes it possible to create more accurate machine-learning classifiers. A human labeler can generate difficulty scores in two ways: sitting in the back seat of the vehicle during the experiment or a graphical user interface for *post hoc* analysis. For *post hoc* labeling, we constructed a user interface that displays a video recording taken out the front window of the vehicle, such as Fig. 2, and controls that allowed the labeler to move forward and backward in time so that users may

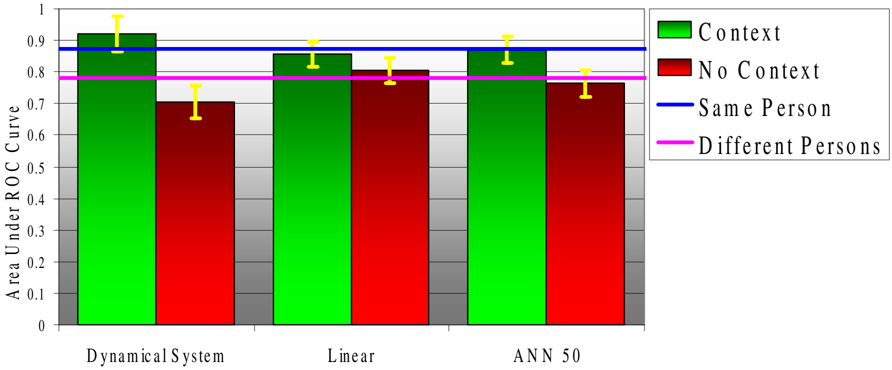


Fig. 4. Performance of the difficulty classifier with and without context recognition in terms of the receiver operator characteristic (ROC) area under curve (AUC). The error bars are the 95% cross-subject confidence intervals. Same persons average an AUC of 0.87 for subsequent scoring of the same round. Different persons agree with each other with an average AUC of 0.78.

adjust difficulty labels to ensure their accuracy. While the human labeler may use the video to generate difficulty labels, the classifier system did not process the images in keeping with the requirement that the system only use sensors currently available on commercial vehicles.

4.2 Off-Road Experiments in Camp Pendleton

We conducted a series of experiments at the United States Marine Corps Base Camp Pendleton, where the experimental platform was a Mercedes-Benz G-class 500 SUV. In these experiments, subjects drove on a mixed semi-improved and off-road circuit four times at 30 km/hour, with each circuit lasting about half an hour. We collected data from nineteen drivers, resulting in 42 hours (609,744 samples) of data. As in our previous experiments, we had to contend with unforeseen events, such as vehicle traffic, road guards, and other equipment. The results described in this paper will be based on the data collected from these experiments (Section 5).

5 Results

To evaluate the results of our driving-difficulty classifier, we compared the context-based difficulty recognizers to those without context recognition. For the results without context recognition, we mapped directly from the regression-coefficients (cf. Fig. 1) to the difficulty labels[1]. In all cases, we tried several regression architectures, including a linear dynamical system, a linear mapping, and a feedforward artificial neural network (ANN).

The linear dynamical system was trained using an iterative one-step optimal Expectation-Maximization routine using least-squares pseudoinversion of the feedforward and feedback matrices. The linear mapping was trained using the closed-form

[1] Mapping from the sensors to driving difficulty did not produce results better than random.

optimal least-squares pseudoinverse. The ANN had *arctan* node activation with 50 hidden units[2], trained with the quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm with Fletcher-type line search [15]. The unsupervised context-learning algorithm was given 7.5 hours (107,313 samples) of unlabeled driving data from which to extract driving context. The supervised-learning algorithms were trained with 4.6 hours (65,678 samples) of labeled driving-difficulty data. The hold-out cross-validation set was 2.4 hours (34,581 samples) of labeled data from subjects not contained in the supervised-training or context-learning sets.

To compare the performance of the different approaches, we use the correlation coefficient between the estimated driving difficulty (the scalar 1-100 values) and the ground-truth driving difficulty generated by the human labelers. To baseline the results, we also asked human labelers to generate difficulty labels for the same round on subsequent days, and asked different human labelers to generate difficulty labels, and compared their results to other labelers. This yields a correlation for how consistent humans are with themselves, and how consistent different persons are with each other. The results are summarized in Fig. 3. In terms of correlation coefficients, all context-based difficulty classifiers outperform those that do not use context recognition. The context-based linear dynamical systems (ρ=0.76) and the context-based ANN (ρ=0.73) perform to the consistency level of different persons with each other (ρ=0.74). The best non-context-based classifier, the ANN, achieved a statistically significantly worse correlation of ρ=0.56.

Another measure of performance is the receiver operating characteristic (ROC) area under curve (AUC) measure [3]. In our case, this measures the probability that a difficulty estimate will agree with a ground-truth label that the situation is "too difficult," cf. Section 4.1.2. The results are summarized in Fig. 4. Once again, all context-based classifiers outperform those that do not use context recognition. The best performer was the context-based linear dynamical system (AUC=0.92), which performed as well as the self-consistency of human labelers (AUC=0.87). The best non-context-based classifier, the linear mapping, achieved a statistically significantly worse result of AUC=0.80. Thus, the best context-based classifier reduces the AUC error rate by almost 60% over those classifiers that do not use context recognition, achieving human levels of performance on both correlation and AUC measures.

6 Conclusions and Future Work

We have presented a context-based semi-supervised machine-learning approach to identify difficult driving situations. We showed that context-based classifiers outperform those that do not use context recognition and that a context-based linear dynamical system can achieve human-like performance on real-world experimental data. In future work, we plan to look at techniques for automatically adapting the generalized contexts to the behavior of a new driver. This will create contexts that are representative of the actual person-specific driving style. In addition, because we have much

---

[2] An ANN with 50 hidden units performed better than other hidden-layer sizes on cross-validation data, which is, incidentally, close to the number of contexts discovered by our unsupervised context-learning algorithm on this data set.

more unlabeled data than labeled data, we want to look at bootstrapping techniques for the difficulty scorer.

In the experiments so far, we have applied this technique within the realm of driver overload. In the future, we plan to change our focus to look at the more common condition of driver underload. By underload, we mean those situations that become potentially dangerous because the driver is distracted, inattentive, drowsy, or bored. We plan to extend the context-based approach to unsupervised learning approach in order to identify unusual, potentially dangerous driving situations due to underload.

Acknowledgment

This work was funded by the Office of Naval Research Code 30 and DARPA/DSO/IPTO. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000.

References

1. NHTSA: Traffic safety facts: 2006 data. Technical Report DOT HS 810 807, U.S. Department of Transportation, National Highway Traffic Safety Administration (2007)
2. Klauer, S., Dingus, T., Neale, V., Sudweeks, J., Ramsey, D.: The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data. Technical Report DOT-HS-810-594, National Highway Traffic Safety Administration (2006)
3. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. Wiley-Interscience, New York (2001)
4. Thorpe, C., Hebert, M., Kanade, T., Shafer, S.: Vision and navigation for the Carnegie-Mellon Navlab. IEEE Transactions on Pattern Analysis and Machine Intelligence 10(3), 362–373 (1988)
5. McDowell, K., Nunez, P., Hutchins, S., Metcalfe, J.S.: Secure mobility and the autonomous driver. IEEE Transactions on Robotics 24(3), 688–697 (2008)
6. Montemerlo, M., Thrun, S., Dahlkamp, H., Stavens, D., Strohband, S.: Winning the DARPA Grand Challenge with an AI robot. In: Proceedings of the AAAI National Conference on Artificial Intelligence (2006)
7. Franke, U., Heinrich, S.: Fast obstacle detection for urban traffic situations. IEEE Transactions on Intelligent Transportation Systems 3(3), 173–181 (2002)
8. Fletcher, L., Apostoloff, N., Petersson, L., Zelinsky, A.: Driver assistance systems based on vision in and out of vehicles. In: Proceedings of the IEEE Intelligent Vehicles Symposium (2003)
9. Miller, B.W., Hwang, C.H., Torkkola, K., Massey, N.: An architecture for an intelligent driver assistance system. In: Proceesings of the IEEE Intelligent Vehicles Symposium, pp. 639–644 (2003)
10. Himberg, J., Korpiaho, K., Mannila, H., Tikanmaki, J., Toivonen, H.T.: Time series segmentation for context recognition in mobile devices. In: IEEE International Conference on Data Mining (2001)

11. Boutell, M., Luo, J.: Incorporating temporal context with content for classifying image collections. In: Proceedings of the 17th International Conference on Pattern Recognition, vol. 2, pp. 947–950 (2004)
12. Dixon, K.R., Lippitt, C.E., Forsythe, J.C.: Supervised machine learning for modeling human recognition of vehicle-driving situations. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (2005)
13. Kubat, M., Holte, R.C., Matwin, S.: Machine learning for the detection of oil spills in satellite radar images. Machine Learning 30(2-3), 195–215 (1998)
14. Kohlmorgen, J., Dornhege, G., Braun, M., Blankertz, B., Muller, K.R., Curio, G., Hagemann, K., Bruns, A., Schrauf, M., Kincses, W.: Improving human performance in a real operating environment through real-time mental workload detection. In: Toward Brain-Computer Interfacing, pp. 409–422. MIT Press, Cambridge (2007)
15. Fletcher, R.: Practical Methods of Optimization, 2nd edn. Wiley Interscience, New York (1987)

Neurally-Driven Adaptive Decision Aids

Alexandra Geyer[1], Jared Freeman[1], Denise Nicholson[2], Cali Fidopiastis[2], Phan Luu[3], and Joseph Cohn[4]

[1] Aptima, Inc., 1726 M Street, NW, Suite 900, Washington, DC 20036, USA
{ageyer,jfreeman}@aptima.com
[2] Institute for Simulation & Training, University of Central Florida, Partnership II Building, 3100 Technology Pkwy, Orlando, FL 32826
{dnichols, cfidopia} @ist.ucf.edu
[3] Electrical Geodesics, Inc., Riverfront Research Park, 1600 Millrace Drive, Suite 307, Eugene, OR 97403
pluu@egi.com
[4] Joseph Cohn
navycohn@hotmail.com

Abstract. Warfighters are constantly challenged with increasingly complex mission environments, roles, and tasks, which require rapid and accurate decision making. Most current military and commercial decision aids leverage a single strategy: they retrieve and fuse information about well-defined objects and events for the user. Such aids effectively discourage users from considering contextual information and patterns that may help them recognize or think critically about hostile or innocent events. If a decision aiding system were to be truly effective, its adaptive strategies should be driven by more than manipulation of well-defined information presented to the user. In this paper, we propose several critical factors - (1) Information state, (2) User cognitive state, and (3) Interaction state – that will enable for discern what must be decided and by when; discriminate which cognitive state and process are in play; and assess interactions (queries, selections, etc.) with the information. Most importantly, these factors will allow for a decision aid to capitalize on –the distinctly human ability to find meaning in swarm of objects and events being perceived.

Keywords: adoptive decision aids, intuition, cognitive state, warfighters.

1 Decision Aids

Warfighters are often faced with dynamic and increasingly complex mission environments, roles, and tasks, which necessitate rapid and accurate decision making. Army combat operations occur in densely-populated urban settings where the physical dimensions and cultural and social characteristics of the environment interact to dramatically compress and complicate the dynamics of the battlespace. The enemy is no longer a large, slow-moving, monolithic entity from the Cold War; the enemy is diverse, numerous, and includes asymmetric threats—small bands of unknown and highly adaptive terrorists, insurgents, drug-traffickers, and other criminal elements. There is a need for adaptive decision aids that support operators in these complex,

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 30–34, 2009.

dynamic contexts. Most current military and commercial decision aids leverage a single strategy: they retrieve and fuse information about well-defined objects and events for the user. This is the purpose of track correlation functions (which associate track radar, emissions, ID and other data) on consoles in the AEGIS Combat Information Center (CIC). It is the function of most information management tools for the intelligence community [1]. Such aids effectively discourage users from considering contextual information and patterns that may help them recognize or think critically about hostile or innocent events. These aids, in short, focus user attention on the most recognizable data at the finest level of granularity. They fail to capitalize on – and may even suppress – the distinctly human ability to find meaning in swarm of objects and events they perceive. For example, CIC systems encourage users to focus on the kinematics of individual entities, but do not support inferences about coordinated actions (e.g., one aircraft is the sensor, another is the shooter) between them.

If a decision aiding system were to be truly effective, its adaptive strategies should be driven by more than manipulation of well-defined information presented to the user. An effective system should be able to dynamically adapt the aid it offers as a function of several critical factors: (1) Information state - the evolution of the decision task over time towards a deadline; (2) User cognitive state - the decision maker's cognitive state and decision processes; and (3) Interaction state - the decision maker's interactions (queries, selections, etc.) with the information.

(1) Automated assessment of the state of Information should enable the aid to discern what must be decided and by when. Specifically, these measures enable the system to (A) estimate the focus of decision activity, (B) the time course of a decision (to discriminate early from late decisions, and (C) manipulate that information in many of the ways above (testing, exploration, algorithm support).
(2) Automated assessment of the Cognitive state should enable the aid to discriminate which cognitive state and process are in play.
(3) Automated assessment of Interaction state should provide behavioral data with which to triangulate on information state and cognitive state.

Assessing each of the three states described above presents challenges in an environment in which time is scarce, stakes are high, and uncertainty is ever present. However, the challenge of measuring cognitive state is perhaps the most difficult. Cognitive psychology has developed methods of inferring cognitive state and process from reaction time and accuracy data but these methods are highly artificial, suitable only for laboratories and not for operational environments. However, recent advances in neuroscience may enable us to measure aspects of cognitive state and cognitive process reliably, in operational settings. For example, a number of neuroimaging and neurophisiological studies examined the nature of decision making and the process that underlie it, such as – Intuition [2][3][4]. We are engaged in research in research that focuses on intuition.

2 Intuition

Intuition is often credited with helping warfighters succeed in critical situations. Research in human pattern recognition and decision-making suggest that there is a sixth sense through which humans can sense unique patterns without consciously seeing them [5][6]. This fast-acting mode of recognition may act as a first-pass filter for gleaning insight of an entire scene [7]. Cursory evidence at an aggregate level suggests that this capability, known as “sixth sense” or “intuition” may be detectable at both the behavioral [8] and neurophysiological [2] [3]. We define intuition as an affectively charged, internal cue to the existence of meaningful information in the environment that arises rapidly and unconsciously. Intuition is not a decision or solution. There is accumulating evidence that indicates that this capability may be trainable [5] [8] and malleable [9] in extremely limited contexts, although it is certainly subject to biases [10] and other types of errors [11].

Recent advances in neuroscience may enable us to reliably measure aspects of cognitive states and cognitive processes involved in decision making, in operational settings [2] [3] [4] [12] [13]. For example, Volz et al. (2006) utilized functional Magnetic Resonance Imaging (fMRI) to examine the neural basis of intuition in participants who were engaged in a modified version of the Waterloo Gestalt Closure Task that involved presentation of images that had been fragmented (i.e., some of the pixels were removed) to varying degrees. The participants were instructed to indicate whether each image contained an object. Fragmenting the images as well as the brief presentation (400ms) made it harder for participants to identify objects in the images. Some of the fragmented images were also scrambled in a way that made them appear incoherent. Participants were instructed to use their “feeling” of whether each image contained/did not contain an object, but they did not have to identify the object. The results of this study revealed activations in the median Obitofrontal Cortex (OFC), the lateral portion of the amygdale, anterior insula, and ventral occipito-temporal regions. The authors identified the OFC to be the area subservient for intuitive coherence judgments. Volz et al. (2008) study further confirmed these conclusions as well as demonstrated that activation in the OFC is modality independent.

In order to further characterize the temporal, spatial, and contextual parameters of intuition, we recently carried out an experiment, utilizing high-density array Electroencephalography (EEG). The results of experiment demonstrated activation in the Orbitofrontal Cortex (preciously identified to be associated with intuitive processing, see [3] [4]) as early as 220ms after stimulus onset in response to images that were perceived to be coherent, regardless of whether there was an actual object in the image or not. These results provided additional support for the notion that intuition can be characterized temporally and spatially, and that we can reliably measured its occurrence with neurophysiological tools.

3 Decision Aids Guided by Intuition

The development of reliable measures of intuition provides new opportunities to understand and aid human cognition. Here, speculate on the function of intuition in human cognition, and the way that intuition-aware devices might enhance human performance.

Intuition is a rapid, automatic cue to the decision maker that aspects of the current situation are coherent or meaningful. Intuition in and of itself does not convey the meaning of an object or event. It merely signals that recognition or comprehension of a scene is approaching.

This signal may help a domain expert to manage the decision process. Thus, intuition serves a metacognitive function. For example (see Figure 1):

The presence of intuition indicates to the decision maker that it is better to wait out the process of recognition than to guess, if time is short. Even if time is plentiful, it may be best to wait briefly for recognition to produce solutions instead of (or before) engaging in deliberate analysis.

Conversely, the absence of intuition indicates that waiting for recognition may be futile. Thus, it may be best for the decision maker to apply a default response or guess if time is short; it may be best to invoke deliberate, analytic processes if time is plentiful.

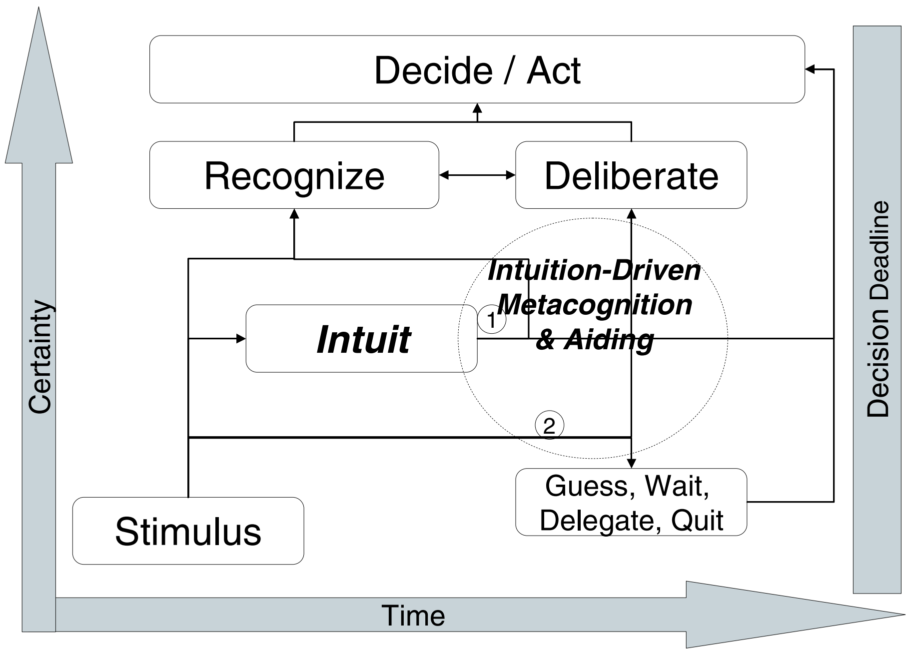


Fig. 1. Intuition may support metacognition and decision aiding. Its presence (1) may promote recognitional processing and rapid action. Its absence (2) may promote guessing or deliberation.

If intuition helps people to manage decision processes and decision time, then reliable measures of intuition might enable us to create a new generation of machines to support decision making.

A decision aid that senses its user's intuitive moments might, perversely, serve its owner best by withholding information that could conflict with the user's rising solution. For example, automated target recognition aid typically overlays sensor imagery (e.g., a vehicle obscured by trees) with a diagrammatic template of the most likely target (e.g., a technical or pickup truck mounted with a machine gun). Given a reliable signal of intuition, the aid might withhold that template while the user's recognition of the scene resolves, with the expectation that the user's response will be more accurate and sufficiently quick.

A decision aid that senses the absence of an intuitive moment might serve its owner by rapidly cueing a solution (e.g., a technical) if time is short. It might, if time is plentiful, present its owner with decision analysis tools (e.g., a decision tree for discriminating technicals from standard pickups, light armored vehicles, etc.).

These are real time, personal applications, in which the intuitions of the decision maker instantaneously drive his (or her) own decision aid.

We are devising an array of other applications that may use measures of intuition in real time to support teams of humans in real time, and other applications that use intuition offline to train autonomous robots.

In sum, reliable measures of intuition may help us to devise a new generation of tools that enhance human decision processes or protect them from interference. The net result should be faster and better decisions when they are needed most.

References

1. Elm, W., Potter, S., Tittle, J., Woods, D., Grossman, J., Patterson, E.: Finding Decision Support Requirements for Effective Intelligence Analysis Tools. In: Human Factors and Ergonomics Society Annual Meeting Proceedings, Cognitive Engineering and Decision Making, vol. (5), pp. 297–301 (2007)
2. Bowden, E.M., Jung-Beeman, M., Fleck, J., Kounios, J.: New Approaches to Demystifying Insight. Trends in Cognitive Sciences 9, 322–328 (2005)
3. Volz, K.G., von Cramon, D.Y.: What Neuroscience Can Tell About Intuitive Processes in the Context of Perceptual Discovery. Journal of Cognitive Neuroscience 18, 2077–2087 (2006)
4. Volz, K.G., Rubsamen, R., von Carmon, D.Y.: Cortical Regions Activated by the Subjective Sense of Perceptual Coherence of Environmental Sounds: A Proposal for a Neuroscience of Intuition. Cognitive, Affective, & Behavioral Neuroscience 8(3), 318–328 (2008)
5. Hogarth, R.M.: Educating Intuition. University of Chicago Press, Chicago (2001)
6. Winerman, L., Kersting, K.: What we know without knowing how. Monitor on Psychology 36, 50–65 (2005)
7. Rensink, R.A.: Visual Sensing Without Seeing. Psychological Science 15(1), 27–32 (2004)
8. Ambady, N., Rosenthal, R.: Half a Minute: Predicting Teacher Evaluations from Thin Slices of Nonverbal Behavior and Physical Attractiveness. Journal of Personality and Social Psychology 64(3), 431–441 (1993)
9. Ambady, N., Gray, H.M.: On Being Sad and Mistaken: Mood Effects on the Accuracy of Thin-Slice Judgments. Journal of Personality & Social Psychology 83(4), 947–961 (2002)
10. Greenwald, A.G., Banaji, M.R., Farnham, S.D., Rudman, L.A., Nosek, B.A., Mellot, D.S.: A Unified Theory of Implicit Attitudes, Stereotypes, Self-concept, and Self-esteem. Psychological review 109, 3–25 (2002)
11. Denes-Raj, V., Epstein, S.: Conflict Between Intuitive and Rational Processing: When People Behave Against Their Better Judgment. Journal of Personality and Social Psychology 66(5), 819–829 (1994)
12. Kounios, J., Frymiare, J.L., Bowden, E.M., Fleck, J.I., Subramaniam, K., Parrish, T.B., et al.: The Prepared Mind: Neural Activity Prior to Problem Presentation Predicts Subsequent Solution by Sudden Insight. Psychological Science 17, 882–890 (2006)
13. Kounios, J., Fleck, J.L., Green, D.L., Payne, L., Stevenson, J.L., Bowden, E.D., Jung-Beeman, M.: The Origins of Insight in Testing-state Brain Activity. Neuropsychologia 46, 281–291 (2008)

Understanding Brain, Cognition, and Behavior in Complex Dynamic Environments

Scott E. Kerick and Kaleb McDowell

US Army Research Laboratory
Aberdeen Proving Ground, MD 21005
{scott.kerick,kaleb.mcdowell}@us.army.mil

Abstract. Many challenges remain for understanding how the human brain functions in complex dynamic environments. For example, how do we measure brain physiology of humans interacting in their natural environments where data acquisition systems are intrusive and environmental and biological artifacts severely confound brain source signals? How do we understand the full context within which the human brain is operating? How do we know which information is most meaningful to extract from the data? How can we best utilize that extracted information and what are the implications for human performance? The papers comprising this section address these questions from conceptual, technical, and applied perspectives. It is clearly seen that significant progress has been made since the inception of the Augmented Cognition program and that, to overcome these challenges, a continued multidisciplinary approach is required across basic and applied research from cognitive scientists, neuroscientists, computer scientists, and engineers.

Keywords: electroencephalography (EEG), natural environment, operational neuroscience, Augmented Cognition, cognitive engineering, human dimension.

1 Introduction

Over the past several decades, the field of neuroscience has made significant contributions to our understanding of human cognition. Neuroimaging, in particular, has unveiled a great deal more about the structure and function of the brain and how mental representations and behavior are generated. Much of this research has been conducted in controlled laboratory environments in which isolated auditory or visual stimuli are presented and simple behavioral responses are required. Moreover, such tightly controlled laboratory settings often study participants in acoustically and eletromagnetically shielded rooms while operating under conditions of minimal or highly restricted movement. Although it has advanced our basic understanding of how the brain functions within highly constrained environments, the extent to which controlled laboratory research generalizes to how the brain functions in complex and dynamic environments in the real world is currently not well understood. In fact, it may be argued that "laboratory studies conceived and interpreted in isolation from real-world experience may do far worse than fail to generalize back to the natural environment; they may *generate* fundamental misunderstandings …" [1; p. 177]. Furthermore, what

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 35–41, 2009.

we know about degeneracy and complexity of biological systems [2,3] suggests that the brain is capable of using its estimated quadrillion neural connections in different ways to accomplish the same task. Based on these concepts, it's likely that there are fundamental differences in how the human brain actually functions to control behavior when it is situated in ecologically valid environments (i.e., situated cognition) relative to that observed in highly controlled laboratory environments.

Because of the complexity of the brain and the natural world, and the inherent measurement challenges of recording neurocognitive activity in uncontrolled environments, we are only now beginning to understand how humans process information and interact in the real world. Ecological approaches [4,5] have for a long time advocated the need to focus on human, task, and environmental interactions to understand behavior in realistic settings. Over the past decade, these concepts have been extended to understanding the interactions between brain functions and operational environments [6-18]. Tools designed to examine these interactions have been and are continuing to rapidly advance through programs such as Augmented Cognition [16-18]. Much progress has been made in neurotechnology as evidenced by advances in sensor technologies [19,20], signal processing techniques such as independent component analysis [21-23], directed component analysis [24], and single-trial phase synchronization [25,26], as well as computational algorithms for classifying cognitive states [13-15,27-29] and brain-computer interfaces [30-32].

While these advancements are enabling preliminary insights into situated cognition, there exists a need to further advance such technologies and validate methodologies for conducting research in real-world environments. Sensor technologies and signal processing techniques have not yet matured to a level at which brain function can be reliably observed in naturalistic settings to the extent possible in laboratory settings [33]. While this goal may ultimately be untenable, science and technology are fast approaching toward this end. Due to challenges of experimental control, one approach is to integrate and synchronize multivariate data (e.g., physiological, behavioral, and contextual) and then apply data mining techniques to search for “hidden” relationships [22,34].

These advancing technologies and methods are expected to provide important insights into how people “think” about the information that they encounter – and how well they can translate that thinking into effective behavior. From an application standpoint, ensuring that people “think well” is non-trivial. The complicated nature of the human-task-environment interactions is seen in the analysis of military and industrial disasters, in which decision makers unsuccessfully interacted with equipment and other personnel in stressful, dynamic environments (e.g., see the shooting down of Iran Air flight 655 by the U.S. Navy in 1988 or the partial core meltdown of the nuclear reactor on Three Mile Island in 1979). Analysis of such disasters reveals that cognitive aspects of complex human-system interactions can have dramatic and unexpected consequences [35]. As the explosive advances in information and computing technologies that have occurred over the past several decades continue, and as the relationships in society become increasingly dynamic and nonlinear, it is expected that the nature of cognitive processing will continue to change from a model that primarily relies on people to one that involves a balance between people and technology.

As a consequence, it is expected that human-system performance in the real world will be largely dependent on how well such systems are cognitively engineered [10,34].

To highlight the state-of-the-art in neurotechnology, the current fundamental research gaps, and the potential benefits of advancing our understanding human cognition in operational environments, we have selected papers in several critical areas. The session starts with Stephen Whitlow presenting recent research on wireless, dry-sensor EEG-based workload classification conducted on dismounted soldiers during performance of military operations in urban terrain (MOUT). This paper frames the problems and illustrates the successes and issues of neurocognitive monitoring of ambulatory soldiers in the real-world. From that basic framework, the next two papers present cutting-edge hardware and software developments for mobile brain imaging. Chin-Teng Lin presents engineering advances of a wireless, dry sensor EEG system featuring micro-electrico-mechanical systems (MEMS) sensors with digital signal processing on a chip. Robert Frank introduces a novel real-time artifact mitigation algorithm based on a spatial filtering to direct the removal of biological artifacts from brain signals in EEG data. The following two papers extend the discussion from EEG to a more comprehensive multidimensional approach to understanding brain and behavior. Scott Makeig presents a new mobile brain/body imaging (MoBI) concept for integrating multisensory inputs such as eye, head, and body movements along with EEG and contextual data from behavior and the environment. Don Tucker then discusses data fusion and data mining approaches to creating and interpreting data sets that include eye- and head-tracking, high-density EEG, and system-based information and the challenges associated with data synchronization and integration. The final two papers in the session present the application of neurotechnology for enhancing our understanding of cognitive states of individuals in ecologically valid task environments. Ruey-Song Huang presents EEG correlates of driving performance based on time-frequency analysis of independent components derived from ICA and discusses implications for the design of human-computer interface design. Bradley Hatfield concludes the session by discussing a broad framework for understanding principles of brain function for highly skilled visuomotor performance. He also presents research on the effects of stress on performance and the application of a neurofeedback training program to enhance performance.

2 Session Papers

1. Whitlow, Mathan, Dorneich: "EEG-based Cognitive Workload Estimation of Mobile Soldiers in Training Missions." One of the most difficult problems facing scientists and engineers is to better understand human dimensions of performance in the real world, especially in complex dynamic environments in which soldiers perform. Whitlow and colleagues confronted this problem head-on by acquiring EEG data continuously from dismounted soldiers during training of military operations in urban environments (MOUT) using a six-channel, wireless, dry electrode EEG system (QUASAR, Inc.) fitted under the helmet. High and load cognitive workload periods were identified from a video log of soldiers performing various tasks throughout training during day and night operations, as rated by independent

observers and by the soldiers themselves immediately following the completion of each training mission, and statistical machine learning techniques were applied to the EEG spectra to determine classification accuracy. Results revealed 75-90% classification accuracy depending on duration of the temporal smoothing windows. Challenges posed by individual differences and dynamically changing tasks are discussed, as are implications for future research.

2. Lin, Ko, Chang, Wang, Chung, Jung: "Wearable & Wireless Brain-Computer Interface and Its Applications." Lin and colleagues introduce a new prototype, four-channel, mobile and wireless EEG system featuring miniature data acquisition circuitry and dry Micro-Electro-Mechanical System (MEMS) electrodes embedded in a headband. The system consists of a data acquisition (DAQ) unit, a wireless-transmission unit, and a real-time signal-processing unit. They also present research from their lab validating the system with participants performing a realistic lane-maintenance driving task in a virtual-reality-based dynamic driving simulator. Results verified that the system performed comparable to established wet-electrode systems. Challenges and future directions for application of this exciting new technology are discussed.
3. Luu, Frank, Kerick, Tucker: "Directed Components Analysis: An Analytic Method for the Removal of Biophysical Artifacts from EEG Data." Luu and colleagues introduce a new signal processing technique, directed components analysis (DCA), for removing biological artifacts from EEG data in real-time. DCA is a spatial filtering method that employs a spatial template to direct the selection of targeted artifacts, is computationally efficient, and can be applied online in real-time. In this paper they examine the effects of undersampling the scalp potential field on the ability of DCA to remove blink artifacts from event-related potential (ERP) data without distortion using high (128 channel) and low (32 channel) density recordings. The results revealed error fractions of .22 and .34 for high and low density recordings, respectively. Strengths and weaknesses of DCA are discussed with respect to alternative methods and future directions are also discussed.
4. Makeig: "Mind Monitoring via Mobile Brain-Body Imaging." Makeig and colleagues expand on existing concepts in brain-computer interface (BCI) design and application based on mobile brain/body imaging (MoBI) for brain/body interface (BBI). MoBI proposes a multisensory modeling approach (brain, eye-movement, body motion and environmental/contextual data integration) to cognitive state monitoring for application to a new, more robust brain/body interface (BBI). This approach extends existing brain-computer interface (BCI) designs by enabling the assessment of complex, natural behaviors in realistic environments and makes greater use of information embedded within the EEG signal (previous BCI systems underutilize information in the EEG; e.g., only one spectral band or time domain signal feature). Additionally, fundamental questions regarding individual differences, brain systems that effect BCI modulation, integration of multisensory inputs, and the effects of training on phasic and tonic brain states are discussed.
5. Tucker, Luu: "Operational Brain Dynamics: Data Fusion Technology for Neurophysiological, Behavioral, and Scenario Context Information in Operational Environments." One major challenge to understanding brain dynamics in operational environments is to be able to synchronize and integrate multiple sources of data from the individual, task, and environment in order to better understand the

operator's current state. Tucker and Luu present a state-of-the-art net-centric, distributed-parallel informatics architecture for increasing the bandwidth of the instrumentation and fused analysis of neurophysiological, behavioral, operational scenario events.

6. Huang, Jung, Makeig: "Tonic Changes in EEG Power Spectra During Simulated Driving." Huang and colleagues present research on the relation between brain activity patterns and driving performance. Independent component analysis (ICA) was applied to EEG data acquired from subjects during a simulated lane-maintenance driving task and time-frequency analysis was conducted on clusters of independent components. The results revealed that several clusters of independent component activities showed tonic elevation in alpha- and theta-band power spectral baseline as reaction time to lane-drift events increased, while other clusters showed broadband or delta-band increases. Implications of this research are discussed with respect to practical applications in human-machine interface/ interaction design.
7. Hatfield, Haufler, Contreras-Vidal: "Brain Processes and Neurofeedback for Performance Enhancement of Precision Motor Behavior." Understanding the how the brain adapts with training (i.e., neural plasticity), how it functions during highly skilled motor behavior, and how stress effects brain function and performance are interesting scientific endeavors with important potential implications for education and training. Hatfield and colleagues discuss a conceptual framework of psychomotor efficiency for motor skill learning and elite-level performance and review current research illustrating the effects of competitive stress on cortical perturbations and shooting performance. They also review new research on how neurofeedback training influences cortical dynamics and shooting performance. Finally, future directions are discussed with respect to the relevance of genetic influences and individual differences in brain function of skilled performers under stress and the role of social factors.

3 Summary

The presentations comprising this session represent advances in basic research and engineering of the hardware and software required for neurocognitive assessment in operational environments, as well as field research, applications and vision for future application. While this session has focused on the ecological principle of examining cognitive function in realistic settings, great progress is being made and will continue to be made through highly-controlled laboratory-based investigation. However, we must focus on converging laboratory and field research with cognitive engineering to ensure the development of system designs that present information to people in ways that enable greater comprehension in shorter durations without inducing undue cognitive demands; intuitive designs that decrease the need for training; and adaptive systems that understand a person's state and adjust training or augment the system accordingly.

Acknowledgements. We would like to thank Henry Girolamo from the Natick Soldier Research, Development, and Engineering Center for his enthusiastic efforts and support in transitioning Augmented Cognition research into the Army. We'd also like to the authors, reviewers, and editors for making this session possible. This document

has been approved for public release in accordance with the Department of Defense; distribution is unlimited.

References

1. Kingstone, A., Smilek, D., Ristic, J., Friesen, C.K., Eastwood, J.D.: Attention Researchers! It is Time to Take a Look at the Real World. Curr. Dir. Psychol. Sci. 12, 176–180 (2003)
2. Edelman, G.M., Gally, J.A.: Degeneracy and Complexity in Biological Systems. Proc. Nat. Acad. Sci. 98, 13763–13768 (2001)
3. Price, C.J., Friston, K.J.: Degeneracy and Cognitive Anatomy. Trends Cog. Sci. 6, 416–421 (2002)
4. Gibson, J.J.: The Theory of Affordances. In: Shaw, R., Bransford, J. (eds.) Perceiving, Acting, and Knowing: Toward an Ecological Psychology, pp. 67–82. Lawrence Erlbaum, Hillsdale (1977)
5. Winkel, G., Saegert, S., Evans, G.W.: An Ecological Perspective on Theory, Methods, and Analysis in Environmental Psychology: Advances and Challenges. J. Environ. Psychol. (in press)
6. Gevins, A., Leong, H., Du, R., Smith, M.E., Le, J., DuRousseau, D., Zhang, J., Libove, J.: Towards Measurement of Brain Function in Operational Environments. Biol. Psychol. 40, 169–186 (1995)
7. Kruse, A.A.: Operational Neuroscience: Neurophysiological Measures in Applied Environments. Aviat. Space Environ. Med. 78(suppl. 1), B191–B194 (2007)
8. Parasuraman, R.: Neuroergonomics: Research and Practice. Theor. Issues Ergon. Sci. 4, 5–20 (2003)
9. Parasuraman, R., Rizzo, M. (eds.): Neuroergonomics: The Brain at Work. University Press, Oxford (2007)
10. Parasuraman, R., Wilson, G.F.: Putting the Brain to Work: Neuroergonomics Past, Present, and Future. Hum. Factors 50, 468–474 (2008)
11. St. John, D., Kobus, D.A., Morrison, J.G., Schmorrow, D.D.: Overview of the DARPA Augmented Cognition Technical Integration Experiment. Int. J. Hum. Comput. Interact. 17, 131–149 (2004)
12. Sterman, M.B., Mann, C.A.: Concepts and Applications in Aviation Performance Evaluation. Biol. Psychol. 40, 115–130 (1995)
13. Wilson, G.F., Fisher, F.: Cognitive Task Classification Based Upon Topographic EEG Data. Biol. Psychol. 40, 239–250 (1995)
14. Wilson, G.F., Russell, C.A.: Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks. Hum. Factors 45, 635–643 (2003)
15. Wilson, G.F., Russell, C.A.: Performance Enhancement in an Uninhabited Air Vehicle Task Using Psychophysiologically Determined Adaptive Aiding. Hum. Factors 49, 1005–1018 (2007)
16. Kruse, A.A., Schmorrow, D.D.: Session Overview: Foundations of Augmented Cognition. In: Schmorrow, D.D. (ed.) Foundations of Augmented Cognition, pp. 441–445. Lawrence Earlbaum Associates, Mawah (2005)
17. Schmorrow, D.D., Reeves, L.M. (eds.): Foundations of Augmented Cognition. Springer, Heidelberg (2007)
18. Schmorrow, D.D. (ed.): Foundations of Augmented Cognition. Taylor and Francis, Inc., New York (2005)

19. Izzetoglu, M., Izzetoglu, K., Bunce, S., Ayaz, H., Devaraj, A., Onaral, B., Pourrezaei, K.: Functional Near-Infrared Neuroimaging. IEEE Trans. Neural Syst. Rehabil. Eng. 13, 153–159 (2005)
20. Matthews, R., McDonald, N.J., Anumula, H., Woodward, J., Turner, P.J., Steindorf, M.A., Chang, K., Pendelton, J.M.: Novel Hybrid Bioelectrodes for Ambulatory Zero-Prep EEG Measurements Using Multi-Channel Wireless EEG System. In: Schmorrow, D.D., Reeves, L.M. (eds.) HCII 2007 and FAC 2007. LNCS (LNAI), vol. 4565, pp. 137–146. Springer, Heidelberg (2007)
21. Jung, T.P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., Sejnowski, T.J.: Analysis and Visualization of Single-Trial Event-Related Potentials. Hum. Brain Mapp. 14, 166–185 (2001)
22. Makeig, S., Debener, S., Onton, J., Delorme, A.: Mining Event-Related Brain Dynamics. Trends Cog. Sci. 8, 204–210 (2004)
23. Onton, J., Westerfield, M., Townsend, J., Makeig, S.: Imaging Human EEG Dynamics Using Independent Component Analysis. Neurosci. Biobehav. Rev. 30, 808–822 (2006)
24. Poolman, P., Frank, R.M., Luu, P., Pederson, S.M., Tucker, D.M.: A Single-Trial Analytic Framework for EEG Analysis and its Application to Target Detection and Classification. Neuroimage 42, 787–798 (2008)
25. Lachaux, J.P., Rodriguez, E., Van Quyen, M.L., Lutz, A., Martinerie, J., Varela, F.: Studying Single-Trials of Phase Synchronous Activity in the Brain. Int. J. of Bifurc. Chaos 10, 2429–2439 (2000)
26. Varela, F., Lachaux, J.P., Rodriguez, E., Martinerie, J.: The Brainweb: Phase Synchronization and Large-Scale Integration. Nat. Rev. Neurosci. 2, 229–239 (2001)
27. Berka, C., Levendowski, D.J., Cvetinovic, M.M., Petrovic, M.M., Davis, G., Lumicao, M.N., Zivkovic, V.T., Popovic, M.V., Olmstead, R.: Real-Time Analysis of EEG Indexes of Alertness, Cognition, and Memory Acquired with a Wireless EEG Headset. Int. J. Hum. Comput. Interact. 17, 151–170 (2004)
28. Dorneich, M.C., Whitlow, S.D., Mathan, S., Carciofini, J., Ververs, P.M.: The Communications Scheduler: A Task Scheduling Mitigation for a Closed-Loop Adaptive System. In: Schmorrow, D.D. (ed.) Foundations of Augmented Cognition, pp. 132–141. Lawrence Earlbaum Associates, Mawah (2005)
29. Liang, S.F., Lin, C.T., Wu, R.C., Chen, Y.C., Huang, T.Y., Jung, T.P.: Monitoring Driver's Alertness Based on the Driving Performance Estimation and the EEG Power Spectrum Analysis. In: 27th Ann. Int. Conf. Eng. Med. Biol. Soc., vol. 1, pp. 5738–5741 (2005)
30. Birbaumer, N.: Breaking the Silence: Brain-Computer Interfaces (BCI) for Communication and Control. Psychophysiol. 43, 517–532 (2006)
31. Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N., Wolpaw, J.R.: BCI2000: A General-Purpose Brain-Computer Interface (BCI) System. IEEE Trans. Biomed. Engin. 51, 1034–1043 (2004)
32. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain-Computer Interfaces for Communication and Control. Clin. Neurophysiol. 113, 767–791 (2002)
33. Kerick, S.E., Oie, K.S., McDowell, K.: Assessment of EEG Signal Quality in Motion Environments. ARL Technical report (in press)
34. McDowell, K., Oie, K.S., Crabb, B.T., Paul, V., Brunye, T.T.: The Army's Need for Cognitive Engineering. Insight (in press)
35. Cooke, N.J., Durso, F.T.: Stories of Modern Technology Failures and Cognitive Engineering Successes. CRC Press, Boca Raton (2008)

Designing a Control and Visualization System for Off-Highway Machinery According to the Adaptive Automation Paradigm

Stefano Marzani[1], Francesco Tesauri[1], Luca Minin[1], Roberto Montanari[1], and Caterina Calefato[2]

[1] HMI Group, University of Modena and Reggio Emilia, via Amendola 02, 42100 Reggio Emilia (Italy)
[2] Department of Computer Science, University of Turin, C.so Svizzera 185, 10149, Turin (Italy)
{Stefano.Marzani,Francesco.Tesauri,Luca.Minin, Roberto.Montanari}@unimore.it, Caterina.Calefato@di.unito.it

Abstract. This paper aims at describing the requirements of an off-highway human-machine system able to recognize potential risky situations and consequently prevent them. The developed methodology is based on two techniques derived from the field of human factors studies, namely the Hierarchical Task Analysis (HTA) and the Function Allocation (FA), which have been integrated and revised to suit the specific domain of off-highway machinery. The paradigms of adaptive automation and persuasive technology will be followed in the design process. After the off-highway domain analysis a system aiming at improving operator and machine safety is presented. The information system extends the human intelligence monitoring the stability of the machine.

Keywords: Adaptive Automation**,** Collision, Function Allocation, Human-Machine Interaction, Hierarchical Task Analysis, Off-Highway Vehicles, Overun, Rollover, Runover, Safety, Tractors.

1 Introduction

The complexity of on-board equipment for agricultural and off-highway machinery has dramatically grown during the last years. The increased number of functions and devices has led to a substantial modification in working procedures, the contrary of what has been seen in the automotive domain. Cars have been equipped with electronic information- and/or safety-related systems, which have improved the reliability of the vehicles and the user comfort although this has not affected the essential nature of the driving task [1] [2].

On the contrary, the introduction of electronics in agricultural and off-highway machinery has led to a strong centralization of controls inside the vehicle cabin, bringing a significant modification in the way users must manage their working task. This modification has been especially strong in the agricultural field, where most of the tractor functions can be performed from inside the cabin. Moreover, the incoming

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 42–50, 2009.

diffusion of electronics to implements will allow the user to perform settings, calibrations and corrections while inside the tractor. As a consequence, users are facing an increased number of tasks [3]. In spite of its high complexity, the off-highway driving domain has been considered only in a few scientific works concerning human factors and the optimization of the in-vehicle human-machine interaction.

This paper proposes a methodology for selection and subsequent design of tasks that are suitable for partial or total automation within the cabin of an agricultural or off-highway vehicle. The methodology was applied to the domain of agricultural vehicles. The result of the case study led to the development of an on-board application, based on:

- a joystick with haptic force feedback capabilities for a loader of an agricultural machine. The joystick behaved in different ways depending on the scenario around the vehicle.
- a visual display providing relevant vehicle information to the operator. The display alerted the operator of incurring critical conditions, then providing the procedure to keep the system in a safety configuration;
- a set of rules for risk estimation which define the information provision strategies for risk mitigation adopted by the visual display and the joystick.

This work derives from an Italian regionally funded project called PRO-TRACT (www.pro-tract.it).

2 Approach and Method

The methodology that was developed in this study is based on two techniques derived from the field of human factors studies, namely the Hierarchical Task Analysis (HTA) and the Function Allocation (FA), which have been integrated and revised to suit this specific domain.

The design methodology is divided into three steps. In the first step, the most critical tasks has been established and analyzed through the use of a Hierarchical Task Analysis with the decomposition categories presented by [4]. The second step identified the sub-tasks in the vehicle cabin that would be suitable for partial or total automation and subsequently, different alternatives for automation and relevant information visualization have been generated and evaluated. A method for Function Allocation based on the so-called "York-method" [5] was used here. The appliance of the York-method gave appreciable results in a previous work yet, were a forward collision warning system for the automotive domain was analyzed and designed [6]. Moreover these first two steps have been discussed in [7] and here relevant results will be summarized.

In the last step, the control strategies of the joystick and the strategies to provide information on the visual display have been designed with the aim to keep the operator aware of the most frequent and risky accidents highlighted after the first two phases. These accidents are:

- collision;
- overturn.

A risk-focused design methodology has been followed for the development of the control of the joystick and the information provision strategies of the visual display. This methodology has been defined as a set of rules which describe:

- the vehicle information needed to identify the possible risky condition: e.g. vehicle speed, bucket position;
- how the level of risk associated to both accidents is identified according to the value of these information: e.g., high speed and high bucket position might lead to an overturn accident, coded as an *high level* of risk.
- how the identified risk level is translated into a haptic feedback on the joystick and/or a visual information on the visual display.

Risk-based feedback strategies were designed, that prompt users to perform appropriate control actions, by signaling unsafe handling of machine controls and limiting possible actions to a subset of safe actions. Strategies were implemented by exploiting all available channels: information about the machine's status was presented on the display with two color-coded levels of warning (medium-amber, high-red) and icons identifying the kind of risk currently run by the operator (overturn or collision). Finally, a set of icons suggested how the operator could intervene on the on-board controls (pedals, joystick, steering wheel) in order to get the machine back into full safety conditions.

The approach aimed at extending operator's capabilities of monitoring the stability of the machine, which are currently limited on several respects. On existing machinery, visual monitoring is hindered by poor visibility from inside the cabin, often worsened by dust, dirt and high visual load devoted to the working implement; concurrently, auditory monitoring is mainly hindered by engine and implement noise. The haptic channel is hardly employed for monitoring purposes although largely involved in managing the machine controls (i.e. pedals, knobs, joysticks). Haptic channels has a strong potential for information provision, as operators perform several safety-critical actions (i.e.: controlling implements) by using physical devices: this qualifies them as strong candidates for delivering safety critical information, and for limiting possible actions when such limitation may prevent severe consequences. As a whole, operators' situation awareness is improved, and capability of avoiding risky operations is boosted.

3 The Off Highway Domain

The off-highway domain encompasses farm machinery, construction machinery and special vehicles. The operative context of such vehicles is heterogeneous, therefore operators often accomplish very different tasks. This heterogeneity is due to several aspects:

1. the environment: visibility, climate
2. working area: type of farming and used machinery, road transit
3. equipment: tools layout
4. operator posture: slopes and equipment position

5. telemanipulation: correct action, force, weight perception, consequences on the material integrity
6. precision Farming[1]: activity organisation

In the automotive domain, the driver should lead the vehicle along the road, keeping it safely; the human machine systems complete the driving task and give driver further information. In the off-highway domain instead, the vehicle transition path is not clearly defined and it lacks of precise road signals. Hence the human machine systems are strictly related to the working task, whose are often the exclusive control system.

Recently, some ISO norms (i.e. ISO 11783) have take care of various on-board devices, like virtual terminals, aiming at define physical constraints. These norms are not enough to deal with the off-highway domain complexity because they consider only technological and productive aspects, which represent a little set of the available vehicle functions. It is needed, in the future, the definition of "Off-highway performance science", like the automotive "Driving Science" [8].

Nevertheless there are some user-centred design principles feasible for this domain:

1. **Display**: displays are classified by i) physical features; ii) target user features; iii) task features. The designer aim is to obtain a correct mapping between the display shape and layout and the task to be accomplished, taking care of strength and weakness of human perception, attention, cognition memory, and mental model [9].
2. **Control**: control involves two crucial processes: i) user action selection and performance; ii) feedback loop[2]. Selecting an action is a process influenced by several factors [10] [11] i) decision complexity [12] [13]; ii) events expectation [9]; iv) compatibility between stimulus and response [14], v) trade-off between speed and accuracy; vi) feedback (feedback is dramatically relevant in the joystick design, because in drive-by-wire vehicles the mechanical feedback are missing and they should be emulated by the multifunction joystick).
3. **automation**: automation is generally applied in order to ensure efficiency or to perform dangerous or heavy tasks. For example in the off-highway domain automation is applied in the headline turn function[3].

3.1 On-Board Equipment: Technology Evolution and Safety

The complex growth of on-board equipment for farm tractors and other off-highway machinery has been rather disordered in terms of its consequences for vehicle dashboard and cabin design, into which many complex human-machine interfaces have been introduced [15].

---

1 Precision Farming refers to the in-field variability, By using satellite data to determine soil conditions and plant development, these technologies can lower the production cost by fine-tuning seeding, fertilizer, chemical and water use, and potentially increasing production and lowering costs

2 The feedback loop allows the user to evaluate whether his/her action obtained the desired effects.

3 This automated system optimize all the operations needed to set the vehicle to work on a new field section, whenever the last one is finished.

The introduction in the automotive domain of new electronic devices (both informatics and driving assistance systems) that improve the reliability of the vehicle and the user's comfort has not altered the essential nature of the automobile driving task [2] [1], leading to a proliferation of controls in the cabin, which significantly modifies the way users must manage their working task. The use of electronics to manage implements (hitches, loaders, bailers, sprayers, etc.) allows the user to perform settings, calibrations, and corrections while inside the tractor [7]. As a consequence, users must perform an increasing number of tasks [3]. This means that the evolution trend is technological-centred [16]: there is little attention for the consequences innovations have on the operator-vehicle system.

On one hand, automation improves efficiency, on the other hand it dramatically impacts on the human performance. The monitoring of an automated task puts the operator "out of the control loop" [17], favouring his/her deskilling [18], which has crucial consequences in dealing with malfunctions and high risk situations. The deskilling effects would make safety and efficiency worse.

The final aim of a human-centered automation system is the operator and machine safety, putting to use the adaptive automation and the persuasive technologies paradigms [19].

Accident data collected in the last few years (National Ag Safety Database [NASD], 2003) identify the task of moving loads with front-loading tractors as one of the riskiest operations in the field. Furthermore, lifting and moving operations conducted with front loaders are related to rollover and runover events [7].

4 The Riskiest Task: An Overview and a Method of Analysis

The methodology proposed in this article had been implemented during the development of a risk mitigation system for farm tractors as part of the Italian publicly funded project Pro-Tract[4].

During the Pro-Tract project in-depth interviews with experienced drivers were conducted. Drivers said that avoiding a rollover was their most serious safety concern. After the goal was defined, the Hierarchical Task Analysis (HTA) was carried out on the basis of detailed information from interviews and video recordings of seven tractor operators performing this task [7].

The HTA enabled designers to isolate criticalities in tasks carried out by operators inside the cab and, consequently, the automation designs could be targeted for solving very specific problems (e.g., operator overload). The function allocation (FA) provided criteria for assessing the feasibility of viable design alternatives, which facilitated this phase and reduced the potential randomness of the selection process. On the human factors side, this method helped developing detailed descriptions of farmers' activities during fieldwork. Differences between off-highway machinery and the automobile-rooted concept of "driving" (starting from the user's posture inside the

[4] The partners were Comer Industries S.p.A., a mechatronic research and manufacturing company; Ognibene S.p.A., a hydraulic manufacturer; Walvoil S.p.A., a manufacturer of hydraulic components and joysticks; and the Human Machine Interaction Group at the University of Modena and Reggio Emilia, Italy.

cab) could be highlighted. The apparent similarities between automotive driving and off-highway driving were found to be inaccurate [7].

In next paragraph, the methodology to identify requirements for warning systems will be proposed. The warnings aimed at reducing the risk of rollover and runover accidents while moving loads (e.g., hay, grain, soil) with front-loading tractors.

4.1 Application of Hierarchical Task Analysis and of Function Allocation to the Riskiest Task

The HTA [4] [20] allowed at creating a precise description of the tasks being considered. The analysis consisted of decomposing the tasks into elementary units and organizing them into three hierarchical levels:

1. goals (system state to be achieved),
2. tasks (structured sets of activities required for achieving goals), and
3. subtasks or operations (different sequences of actions that the machine-operator system must perform).

The HTA was applied to one of the most risky work situations for operators of farm tractors, described as follows: “To move a load (e.g., equipment, hay, sand, manure) by means of an implement connected to a farm tractor (operator in cabin with engine on).” Rollovers and pedestrian runovers are two of the most serious and frequent accident types with farm tractors [21] during this activity,

Hence, this task was decomposed in sub-task, in turn decomposed in other subtasks, each of them performed with specific modalities. The template in which the descriptions were included was based on the table proposed by [4]. From the HTA arose one key tasks: an example is “Check physical obstacles”, which may put both the operator and pedestrians in imminent danger. The following function analysis (FA) showed that this specific subtask could be partially automated [7]. The tasks structured description derived from the decomposition analysis facilitated the scenario description required for the FA based on the York-method [5], in which different scenarios and trade-offs are used with regard to the allocation of tasks and operations between humans and machines. Task analysis and function allocation both aim to match the human abilities with the system ones, in order to automate the tasks best suited to machines and to maintain as manual the functions best suited to human [22], considering also the benefits with respect to workload, performance, and situation awareness.

Once the basilar functions have been founded, they will be allocated to the machine or to the operator, considering that “a function may be separable from all roles, and technically feasible and cost effective to automate, in which case the function may be totally automated. Alternatively it is possible that the function maps entirely to one of the roles, and is infeasible to automate, in which case the function is totally performed within that role. In most cases however functions fit into neither category. In this situation the function is to be partially automated” [22], [7].

4.2 Constructing Scenarios and Evaluating Candidates for Total Automation

The scenarios were selected in order to focus on critical situations (for example, where workload is likely to be high). Each scenario was described with a subset of functions required in the scenario. In our case, the HTA and decomposition categories defined the most critical subtasks within the areas associated with rollover and runover risks. Hence, we selected the scenarios to include those two risky areas. Each scenario was then described with a subset of functions that were evaluated as candidates for total automation.

Tasks that have low significance for the operator's role are generally proposed as candidates for total automation. For example, for the "work in courtyard" scenario, the operator role is: "the operator is the only person responsible for an efficient (quickness and accuracy) and safe execution of the job. During execution, he/she must prioritize safety over efficiency. Safety is prioritized according to the following hierarchy: safety of vulnerable persons (e.g., pedestrians), own personal safety, safety of objects and items. During the execution of the task, he/she monitors the protection of machines, vehicles, buildings, and the transported load". The function "estimation of the terrain slope" could be totally automated, while "keep the implement height under the critical threshold" is an example of a partially automated task, that is done by the operator with assistance from the vehicle, which provides information about load height or prevents the operator from raising the implement too high for safe operation [7].

When automation would decrease the operator's overall performance it means that these functions should not be fully automated, either because of a substantial cost or because they were considered to be central to the operator's role [7].

4.3 Development of the Risk Mitigation System

The final phase of the work consisted of defining how to translate the inputs from the machine (e.g., terrain slope) or the operator (e.g., a sudden turn) into the in-cab devices. The solution advanced by the PRO-TRACT project team consisted of an on-board computer screen and a haptic joystick. These devices should be set and used in order to allow the operator to safely perform critical task.

When the risk is high, a suggestion is displayed on the screen about how the operator can mitigate the risk. The suggestion is given in descriptive symbols and icons with color coding that corresponds to the risk level and is accompanied by a brief text message. Simultaneously, a beep sounds to draw the operator's attention to the display. For instance, in the case of a high risk for rollover, the display suggestion would most likely be to lower the loader. Simultaneously, a partial automation device would intervene and impede certain joystick movements by generating resistance in the direction that would raise the loader to an unsafe height. On the other hand, if there is less of a risk for rollover, only the visual and auditory warnings would be presented, leaving the joystick functions unaffected.

This solution would require a low R&D cost, given that the main barrier is the calculation of the risk level for rollover or runover. Joysticks for haptic feedback exist on the market; the display design doesn't require significant costs, and the risk parameters would be measured by sensors installed in the vehicle. Several sensor packages that would meet this need have been developed for the automobile market.

4.3.1 The on Board Computer Screen

This display is foreseen in the ISO11783 norm. It inform the operator about the status and the layout of the vehicle equipment. The display a has a data mask and soft keys. The display must allow the operator to build the correct scenario risk evaluation, giving him/her useful hints to mitigate the risk.

4.3.2 The HapticJoystick

The joystick has four motion possibilities (up/down, left/right) and some buttons. It is placed on the operator's right side. The joystick palys a crucial role, because it is the control device of the vehicle equipment and in the most part of dangerous situations it is the decisive intervention tool. Its haptic feedback allows the operator which are the correct or inadequate manoeuvre.

5 Conclusion

Thanks to the hierarchical task analysis and the function allocation methodologies operative scenario were traced and matched with design solutions adequate to the detected risk level (rollover vs pedestrian runover) and to the working context (courtyard vs field). Thanks to the function allocation activity, designers have at disposal a general description of the system behaviour, essential to write the system requirements.

The proposed methodology allowed us at selecting tasks suitable for partial or total automation concerning off-highway vehicle maneuvers.

The resulting on-board application and risk mitigation system, were formed by:

- a haptic joystick with force feedback that behaved in different ways depending on the current scenario.
- a visual display providing crucial vehicle information to the operator.
- a set of rules for risk estimation which define the information provision strategies adopted by the visual display and the joystick.

References

1. Michon, J.A.: Generic Intelligent Driver Support. Taylor and Francis, London (1993)
2. Hoedemaeker, M., De Ridder, S.N., Janssen, W.H.: Review of European Human Factors Research on Adaptive Interface Technologies for Automobiles. TNO Human Factors, Soesterberg, The Netherlands, TM-02_C031 (2002)
3. Fuller, R.: Towards a general theory of driver behaviour. Accident Analysis & Prevention 37(3), 461–472 (2005)
4. Kirwan, Ainsworth, L.K. (eds.): A guide to task analysis. Taylor & Francis, London (1992)
5. Harrison, D.M.D., Wright, P.C.: Allocation of function: scenarios, context and the economics of effort. International Journal of Human-Computer Studies 52, 289–318 (2000)
6. Calefato, C., Montanari, R., Tango, F.: Advanced Drivers Assistant Systems in automation. In: 12th International Conference on Human-Computer Interaction HCI International 2007, Beijing (China), July 22-27 (2007)

7. Etzler, L., Marzani, S., Montanari, R., Tesauri, F.: Risk mitigating on-board equipment for farm tractors - Hierarchical Task Analysis and Function Allocation to design warning systems aimed at reducing rollover and run-over accidents. Ergonomics in Design, the Quarterly of Human Factors Applications (2007)
8. Green, P.: Driver Distraction, Telematics Design, and Workload Managers: Safety Issues and Solutions. SAE Paper Number 2004-21-0022 (2004)
9. Wickens, C.D., Lee, J., Liu, Y., Gordon-Becker: Introduction to Human Factors Engineering. Pearson Prentice Hall, NJ (2003)
10. Fitts, P.M., Posner, M.I.: Human Performance. Brooks and Cole, Monterey (1967)
11. Wickens, C.D., Hollands, J.G.: Engineering Psychology and Human Performance. Prentice Hall, NJ (2000)
12. Hick, W.E.: On the rate of gain of information. Quarterly Journal of Experimental Psychology 4, 11–26 (1952)
13. Hyman, R.: Stimulus information as a determinant of reaction time. Journal of Experimental Psychology 45, 188–196 (1953)
14. Fitts, P.M., Seeger, C.M.: S-R compatibility: spatial characteristics of stimulus and re-sponse codes. Journal of Experimental Psychology 46, 199–210 (1953)
15. Bellotti, F., De Gloria, A., Montanari, R., Dosio, N., Morreale, D.: COMUNICAR: Designing a multimedia, context-aware human-machine interface for cars. Cognition, Technology & Work 7(1), 36–45 (2005)
16. Billings, C.E.: Aviation automation: the search for a human- centered approach. Lawrence Erlbaum Associates, Hillsdale (1996)
17. Endsley, M.R., Kiris, E.O.: The Out-of-the-Loop Performance Problem and Level of Control in Automation. Human Factors 37(2), 381–394 (1995)
18. Prinzel, Lawrence J.: Team-Centerede perspective for Adaptive Automation Design, Langley Research Center, Hampton, Virginia, available from NASA Center for Areospace In-formation (CASI) (2003)
19. Fogg, B.J.: Persuasive technologies: Using Computers to Change What We Think and Do. Morgan Kaufmann Publishers, San Francisco (2003)
20. Stanton, N.A.: Hierarchical task analysis: developments, applications and extensions. Applied Ergonomics 37, 55–79 (2006); invited paper to special issue on Fundamental Reviews of Ergonomics
21. National Ag Safety Database (2003), `http://www.cdc.gov/nasd/`
22. Harrison, M.D., Johnson, P.D., Wright, P.C.: Automating Function in multi agent control systems: supporting the decision process. In: Proceedings of the Tenth Safety-critical Systems Symposium (2002)

Context-Dependent Force-Feedback Steering Wheel to Enhance Drivers' On-Road Performances

Luca Minin[1], Stefano Marzani[1], Francesco Tesauri[1], Roberto Montanari[1], and Caterina Calefato[2]

[1] HMI Group, University of Modena and Reggio Emilia, via Amendola 02, 42100 Reggio Emilia (Italy)
[2] Department of Computer Science, University of Turin, C.so Svizzera 185, 10149, Turin (Italy)
{Luca.Minin,Stefano.Marzani,Francesco.Tesauri, Roberto.Montanari}@unimore.it, Caterina.Calefato@di.unito.it

Abstract. In this paper the topic of the augmented cognition applied to the driving task, and specifically to the steering maneuver, is discussed. We analyze how the presence of haptic feedback on the steering wheel could help drivers to perform a visually-guided task by providing relevant information like vehicle speed and trajectory. Starting from these considerations, a Context-Dependant Steering Wheel force feedback (CDSW) had been developed, able to provide to the driver the most suitable feeling of the vehicle dynamics according to the driven context. With a driving simulator the CSWD software had been tested twice and then compared with a traditional steering wheel.

Keywords: adaptive steering wheel, augmented cognition, driver performance, force feedback, haptic feedback, lane change.

1 Introduction

Driving a car is usually considered a visual-guided task, where drivers are asked to couple their visual perception of the context with actions on the vehicle trajectory.

In the case of steering, drivers should translate their visual perception of vehicle speed, position, road geometries and presence of obstacles in steering wheel movements. For a decade researchers [1] have been finding that the presence of haptic feedback on the steering wheel could help drivers to perform a visually-guided task by providing relevant information like vehicle speed and trajectory. Referring to the augmented cognition field, we can assess that when using a haptic assist steering wheel rather than a traditional passive steering wheel, drivers are better able to follow a reference path and at the same time, they required fewer visual cues [2].

The effect of steering torque and steering gain, respectively, on the driver's feeling, can be investigated by using some objective measurement parameters, as showed in [3]. The researchers highlight three main conclusions [3]:

1. "The driver's good feeling is influenced by two factors, namely the steering torque magnitude and steering torque delay to driver's steer input. Firstly, steering torque

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 51–57, 2009.

magnitude is related to driver's physical effort, and its optimum value depends on driver's individual physical capability. Secondly, steering torque delay is related to driver's mental effort, where minimum delay is desirable by both drivers. Driver's driving manner is considered to have an influence on driver's ability to adapt to larger steering torque delay".
2. "Higher steering gain (smaller overall steering gear ratio) reduces the driver's physical effort in terms of amount of steer angle and steer velocity required to complete the lane change task, and thus contributes to driver's ease of driving".
3. "It is assumed that the amount of steering torque desired by the driver remains the same for a particular driving task".

Many studies showed also how performance varied as a function of the type of road used in the test (e.g. highway, rural road, urban road – [4]) or the specific driving task (e.g. car following, lane change - [5]).

In more recent years, other studies found that drivers performance could be affected by the specific type of road or force feedback reproduced by the steering wheel ([6] [7] [8] [9]. This suggests that performances could be improved by reproducing on the steering wheel the most suitable force feedback for a specific driving context. Some studies showed how adding vehicle behavior feedback (e.g. lateral acceleration, yaw rate) on the steering wheel provides steering reactivity torque with maneuverability superior to that of a conventional control [7] [8].

During one of our latest research [10], six force feedback had been compared among three different driving contexts. Results suggested that the effect of the type of force feedback in terms of drivers' performance could depend on the specific context.

Starting from these findings we developed a Context-Dependant Steering Wheel force feedback (CDSW), able to provide to the driver the most suitable feeling of the vehicle dynamics according to the driven context. We reproduced CDSW software on the haptic steering wheel of a driving simulator; finally, we tested it to evaluate the effects on drivers' performances compared with a traditional steering wheel.

The experiment conducted in [10] has been repeated in a more recent report [11], whose main results will be explained in this paper. In fact, during the last experiment authors found that for each context, at least one force feedback algorithm allowed drivers to better perform the driving task.

2 Selection of the Set of Best Force-Feedback Algorithms

The force-feedback algorithms used for the development of the CDSW and the correspondent scenario where they are activated had been selected from a previous study [11].

In this work a driving experiment has been conducted on eighteen participants, divided into six groups (3 participants/group) and each group was asked to drive using a specific force-feedback reproduced by a haptic steering wheel.

- Three force-feedback based on a steering angle (two linear and one parabolic simulating the self aligning moment of the tyre);
- three based on the vehicle dynamic (speed, lateral acceleration and yaw rate).
  Three different scenarios were used: urban road, rural road and freeway.

Force-feedback were reproduced by a haptic steering wheel installed on a driving simulator. By means of HFC (High Frequency Component of steering angle – [12] drivers' performance has been evaluated

The ANOVA (ANalysis of VAriance) with scenario as within-subject variable and FF dynamic as between-subjects variable showed a main effect of scenario ($p<.0.001$). The effect of force-feedback significantly interacted with scenario ($p<0.001$). In particular, lower values of HCF have been founded in the following conditions:

- freeway: force feedback proportional to tyres self alignment force, simulated with a proportional to steering angle function with a coefficient of 0.052 deg/Nm;
- urban road: like highway but with a coefficient of 0.285 deg/Nm;
- rural road: proportional to vehicle lateral acceleration.

3 Context Dependant Steering Wheel Design

We implemented the best performing algorithms into our Context-Dependant Steering Wheel force feedback (CDSW) software. We reproduced the CDSW with Matlab Simulink and Stateflow[1] because this software could be easily interfaced with the haptic steering wheel of the driving simulator we used to test its efficacy.

Information on vehicle dynamic needed to compute each algorithm were available on the software of the simulator, as well as data needed to recognized the context.

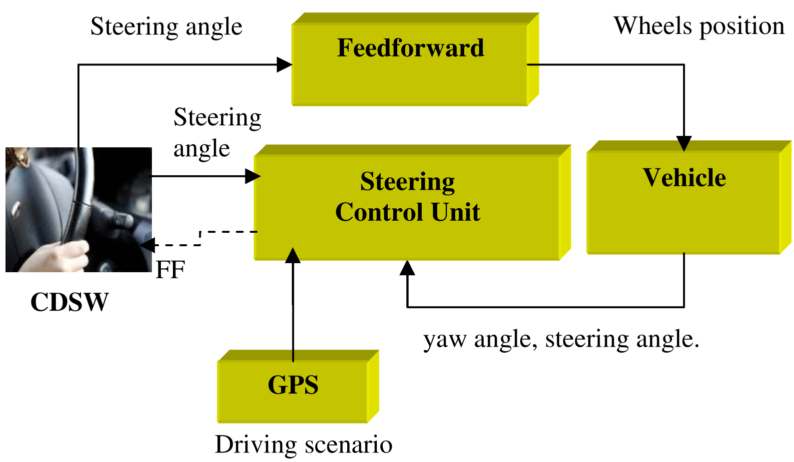


Fig. 1. CDSW implemented on the driving simulator

While driving, a GPS (or the most suitable navigation system) identifies the scenario (e.g. freeway); then, according to the type of FF and scenario, a steering control unit selects and reproduces the pre-loaded best-performing FF. This solution could be easily implemented on a Steer-By-Wire (SBW) system, that is a steering wheel

[1] www.mathworks.com

electronically-controlled, and without mechanical junctions between the front wheels and the steering control.

The CDSW also includes a smooth transition function (see the bottom of the figure below) from a force feedback algorithm (FF1, FF2, FF3) to the next expected for the recognized context the driver is going to drive. This strategy provides a linear interpolation between two consecutive algorithms which enable the CDSW to move to the next expected force feedback algorithm in a defined distance interval. The strategy has been evaluated by a team of developer and ergonomic experts before implementing it on the final system.

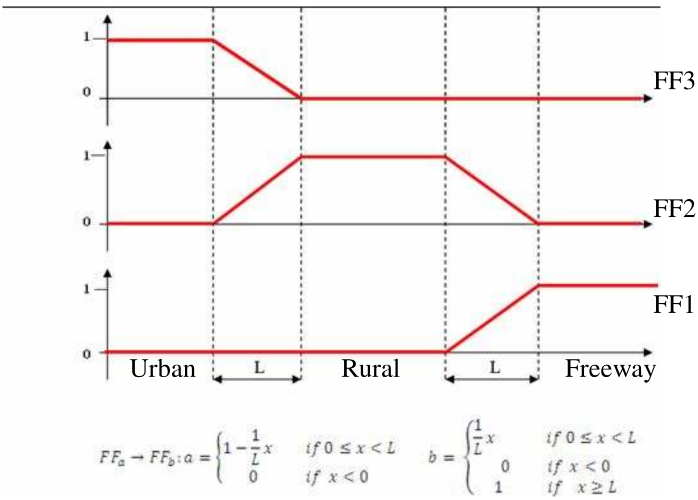


Fig. 2. Interpolation between consecutive Force-Feedback (FF)

The final system represents an adaptive steering wheel that dynamically and smoothly changes the force feedback when the driver move from a context to another. The aim of this device is to improve drivers' performance and perception of the driving task, by providing the most relevant information on vehicle dynamic drivers need in a specific context.

4 Context Dependant Steering Wheel Test on Driving Simulator

The purpose of the previous test was to evaluate whether driving performance could be improved by six different-steering Force-Feedback (FF) dynamics in different driving scenarios.

The test described in this work aimed to assess the performances of the drivers with and without the use of the CDSW in a driving context that changes, in terms of road scenario.

Our objective was to confirm that providing a force-feedback based on information from vehicle dynamic (i.e. angle of steering and lateral acceleration) can improve drivers' behaviour in the lateral control of the car, taking into account. As a recent study

[2] showed by a user test on a fixed-base driving simulator, a path following task can be improved by using a motorized steering wheel. The test results indicated that the haptic assist through the steering wheel improves lane keeping by at least 30% reduces visual demand by 29% ($p<0.0001$) and improves reaction time by 18 ms ($p=0.0009$).

4.1 Study Protocol

Twelve participants in the age between 22 and 42 years old with a 4 year driving experience have been involved in the preliminary evaluation of the CDSW efficacy in a test driving .

We asked drivers to drive for 15 minutes in a scenario with three consecutive driving context, respectively freeway, rural and urban road. Six drivers drove with a traditional steering wheel force feedback (control group); the other six drove with the CDSW system.

Performance has been evaluated by mean of indicators previously used in [6], [10], [11], based on vehicle lateral deviation and steering wheel angle variability, namely:

- Standard deviation of Steering Angle;
- Standard deviation of Time To line Crossing [12];
- High Frequency Component of steering angle.

Data needed to the computation of each indicator have been recorded by the simulator during the experiment at a frequency of 20Hz.

4.2 Experiment Results

Driver performance was submitted to an analysis of variance (ANOVA) with scenario as within-subject variable and drivers' group as between-subjects variable.

The analysis confirmed the main effect on scenario as showed in [11] for all the indicators, with a significance different for each indicator. Furthermore, it was also found a significant difference between the two groups, specifically pointing in favor of the CDSW. The following table shows the results concerning Standard Deviation of steering angle.

Table 1. p-values of the Standard Deviation of steering angle, with reference to the scenario and drivers' group

P - scenario	*P – drivers' group*
0.0194	0.0106

The following figure shows the observed mean values of Standard Deviation of steering angle of each participant belonging to the different groups.

The left side of the figure represents the between-subjects analysis conducted with reference to the group factor, while in the right side the interaction between group and scenario is considered. Both conditions highlight the differences in performance driving with and without the CDSW.

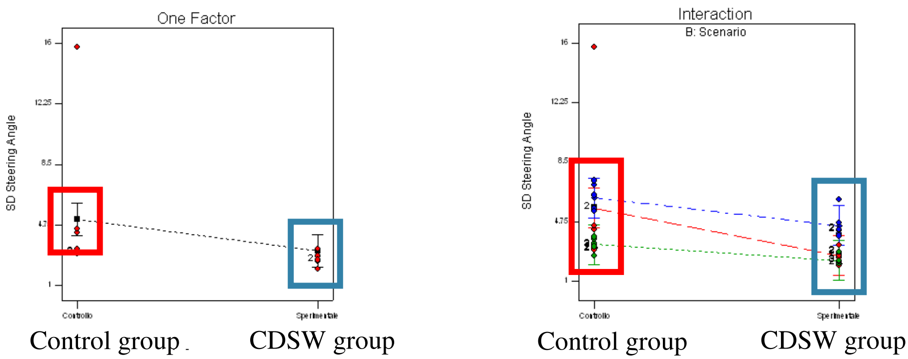


Fig. 3. Drivers' group performances measured with Standard Deviation of steering angle

5 Conclusions

Both the test had confirmed the hypothesis of this work: the driving performance significantly improved when the system activated the FF model. The model is particularly effective in the urban road. These results had been evaluated considering also that the transition strategy among force-feedback models was not intrusive in the vehicle control.

Evaluating the results of our test referring to the augmented cognition field, we may agree that observations confirm the essential role of coherent haptic information for driving real cars and simulators, and also suggest the existence of driver adaptation mechanisms in steering control. As other researchers stated, driver behavior adaptation can occur efficiently for a range of steering force feedback configurations, but this range is limited by certain acceptability limits [6].

We interpreted our results comparing them with data arose from other studies in the literature [13] and the conclusions suggested that, using an intelligent, haptic steering wheel rather than a traditional passive steering wheel, drivers are better able to closely follow a reference path while requiring fewer visual cues.

References

1. Bertollini, G.P., Hogan, R.M.: Applying Driving Simulation to Quantify Steering Effort Preference as a Function of Vehicle Speed. Society of Automobile Engineers, 1999-01-0394 (1999)
2. Griffiths, P., Gillespie, R.: Shared Control between Human and Machine: Haptic Display of Automation during Manual Control of Vehicle Heading. In: 12th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, Chicago, IL, USA, pp. 358–366 (2004)
3. Chai, Y.: A Study of Effect of Steering Gain And Steering Torque on Driver's Feeling For SBW Vehicle. In: Proceedings of Fisita world automotive congress, Barcelona (2004)
4. Martens, M.H., van Winsum, W.: Measuring distraction: the Peripheral Detection Task, Internal Report. TNO Human Factors, Soesterberg, The Netherlands (2000)

5. Salvucci, D., Liu, A.: The Time Course of a Lane Change: Driver Control and Eye Movement Behavior. Transportation Research, Part F 5(2), 123–132 (2002)
6. Toffin, D., Reymond, G., Kemeny, A., Droulez, J.: Influence of steering wheel torque feed-back in a dynamic driving simulator. In: Proceedings of the Driving Simulation Conference, Dearborn MI North America, October 8-10 (2003)
7. Kimura, S.: Research on Steering Wheel Control Strategy as Man-Machine Interface for Steer-by-Wire System. In: Proceedings of FISITA world automotive congress, Barcelona, Spain, pp. 29–33 (2005)
8. Mcdonald, W., Hoffman, E.: Review of relationships between steering wheel reversal rate and driving task demand. Human Factors 22, 733–739 (1980)
9. Mourant, R.R., Sadhu, P.: Evaluation of Force Feedback Steering in a Fixed Based Driving Simulator. In: Human Factors and Ergonomics Society 46 th Annual Meeting (2002)
10. Minin, L., Montanari, R., Corbelli, C., Iani, C.: Force Feedback: new frontier as the innovative driving comfort tool. In: Proceedings of the Computer Human Interaction conference (CHI), Florence, Italy (2008)
11. Montanari, R., Minin, L.: Efficacy tests of the Ergonomic Steer-ByWire System, Mectron Deliverable 5.3, Mectron Italian regional project, PRRIITT - 3.4 – Action A, Emilia Romagna Region (2008)
12. Östlund, J., et al.: HASTE Deliverable 1 - Human Machine Interface And the Safety of Traffic in Europe. Project GRD1/2000/25361 (2004)
13. Steele, M., Gillespie, B.: Shared control between human and machine: Using a haptic steering wheel to aid in land vehicle guidance. In: Proceedings of the Human Factors and Ergonomics Society 45th annual meeting, Minneapolis, Minnesota, USA, pp. 1671–1675 (2001)

Where Is My Stuff? Augmenting Finding and Re-finding Information by Spatial Locations and Icon Luminance

J. Michelle Moon and Wai-Tat Fu

Human Factors Division and Beckman Institute
University of Illinois, 1 Airport Road, Savoy, IL 61874
{moon20,wfu}@illinois.edu

AAbstract. We studied how spatial locations and luminance affect finding and re-finding information in a desktop environment. In an experiment conducted with computer icons, fixed locations led to more frequent accesses to icons while change of luminance led to worse recall of icon titles and locations. In an analysis of icon access transition, a sequential search pattern was identified in earlier sessions, which suggests that participants were minimizing efforts in external search and were not utilizing internal memory of titles and locations yet. In later sessions, icon accesses were more focused to information directly relevant to search tasks as participants started using titles and locations for re-finding icons. Results are consistent with the notion that information search behavior is adaptive to the cost-benefit structure of the interface, and search strategies are adaptive to different external representations of icons. Results also suggest that both external representations and human information processes are critical in determining the effectiveness of different GUI designs.

Keywords: Adaptive human behavior, re-finding information, spatial memory, interface design.

1 Introduction

Imagine you are a real estate agent who recommends apartments for your clients. You have a wide range of information (from monthly rent to neighborhood safety) filed on your computer desktop. How would you utilize your distributed information resources (e.g., electronic documents) to make recommendations for different clients across time? Indeed, it is becoming more common that users need to perform routine, comprehensive information search that requires finding, re-finding, and integrating information from multiple sources depending on the dynamic needs of information (e.g. client's requests). Finding and re-finding information becomes a difficult task especially when we don't know where to look for the right information. The current study was motivated by the realization that re-finding information has not been fully studied in the context of computer icon search in a desktop environment.

1.1 Use of Contextual Information in Information Search

Studies suggest that people take advantage of contextual information they have about their information target when looking for electronic information [1, 7]. Especially,

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 58–67, 2009.

information re-finding process relies heavily on the use of contextual information [7]. In GUIs, computer icons not only can be used to associate with information source, but also can convey interpretive meaning with slight alteration of visual features of icons [9, 15]. Therefore, contextual information can be represented as characteristics of interface objects in GUI so that users can easily pick up this information to guide them to find the relevant information. For example, if contextual information such as when and how many times file icons were previously accessed is visually represented (e.g. using varied luminance levels), users may find it easier to re-find information.

Previous findings on icon search process make this claim more convincing. Icon search is known to consist of the two processes: searching the graphic picture of the icon and searching the text label representing the file name [9]. These two processes of icon search suggest the efficacy of combining bottom-up visual information processing and top-down contextual information processing to aid re-finding icons.

1.2 Spatial Memory and Locations of File Icons

Studies [8, 19] suggest that, as a by-product of interaction with interface objects, learning of the locations of these objects is often incidental and effortless. Ehret [8] claimed that location learning is not only pervasive, but also subject to the cost structure of the interface. When search cost was increased, learning and reliance on location knowledge increased as well. However, the efficacy of a spatially oriented approach to object reference in a computing system, whether this approach is used alone or in conjunction with symbolic reference, may be severely impaired [17].

1.3 Luminance Changes and History of Uses

The metaphor of light has been used as a method to implicitly convey information by bridging physical and digital spaces [16]. Luminance of icons seems to be an intuitive feature to visually represent contextual information (history of use). It is less disruptive compared with other features such as color, size, or location of icons which may involve conflicts with long-term memory. In the Windows operating systems, for example, an icon selected becomes highlighted which gives an implicit but intuitive indication that the icon has been just selected. The assumption is that representing history of use by differences in luminance will facilitate re-finding of the icon in the future. However, how exactly this representation may interact with the adaptive processes of the user remains unclear.

1.4 The Theory of Soft Constraints

We hypothesized that changes in external representations will induce adaptive shift in processing strategies. This adaptive perspective casts the interplay between perceptual-motor and memory processes as an optimization process that maximizes the expected utility of the human information processing system by balancing the cost of internal memory retrieval with the cost of external search [10, 11, 12]. Indeed, research has found that people might ignore perfect knowledge in-the-world (KIW) for imperfect knowledge in-the-head (KIH) when information access cost increases, which induces a new set of soft constraints to the dynamic interplays between internal memory and external search processes [13, 14]. The current prediction is that more contextual information will lower the cost of external search, thus encouraging more external search strategies that exploit this interface feature (KIW); and the lack of contextual information will encourage relatively less costly memory-based strategies that rely on internal memory (KIH). An experiment was designed to directly test this prediction.

2 Method

2.1 Participants and Design

64 participants recruited from the University of Illinois community were randomly assigned to one of the four conditions in a 2x2 between-participant design (Table 1). Participants in each condition were given an interface with file icons with or without the luminance and location features. In the Control and Spatial Memory conditions, luminance of icons was identical and did not change throughout the experiment. In the Control condition, locations of file icons were randomly changed after participants opened or closed an icon. In the Spatial Memory condition, locations of file icons were fixed. In the Luminance and Luminance-Spatial Memory conditions, luminance of icons was initially identical. Over time, however, icons accessed more recently or frequently became brighter than the others. In the Luminance condition, locations of file icons were randomly changed, but in the Luminance-Spatial Memory condition, locations of icons were fixed.

Table 1. Experiment Design

	Randomized Location	Fixed Location
No Luminance changes	Control (C)	Spatial Memory (SM)
With Luminance changes	Luminance (L)	Luminance-Spatial Memory (L-SM)

2.2 Task

All participants were given the same set of eight information search tasks. Each participant was instructed to imagine that he or she were a real estate agent, and was asked to recommend an apartment that met different criteria provided by a different client in each of the eight tasks. Participants were instructed to use any information available in the files. There was a time limit of seven minutes for each task. However, accuracy was emphasized over speed in order to encourage more thorough information search. All participants finished the tasks within the time limit.

Search tasks (Fig. 1) were designed such that participants had to make multiple accesses to icons to find, re-find, and integrate information. For example, participants had to view multiple files to figure out the price category (upper, medium, lower) of a certain apartment. Participants were free to access any of the icons in any order as many times as they wanted.

> Sarah just got a pay raise and is looking for the upper range of options. She prefers having a gym or shared pool in the apartment complex. She is an avid collector of Lego and she likes to display big Lego blocks she assembles. Her old apartment was too small for this. Therefore she would prefer more space (more square feet) to use as storage for her Lego collection.

Fig. 1. Example of information search task

After the search tasks, the interface was removed and all participants were given the icon-title-recall test which asked them to write down titles of icons that they saw during the experiment. Then, only participants in the Spatial Memory and the Luminance-Spatial Memory conditions (with fixed-location icons) were also given the icon-location-recall test. Participants were given titles of icons and were asked to recall where exactly each of the icons was located in the interface by filling out each cell in the 4x5 grid (as in Fig. 2) with the matching title.

2.3 Interface

The interface (Fig. 2) simulated a computer desktop with 20 file icons (grey squares in the top left panel) in the black background. Each icon had a short title with the first four letters of the main content of a file (e.g. 'Garf' for a file about 'Garfield apartment'). We used this naming convention to equalize reliance on icon titles to search. The short titles were intended to minimize effects of differences in working memory capacity (longer names would be more taxing for lower-capacity participants). Contents of the files varied from more relevant (e.g. apartment) to less relevant (e.g. neighborhood) to irrelevant information (e.g. football statistics) in terms of how helpful the given information is to find a correct answer. However, participants had to figure out relative importance of files by exploring icons.

When an icon was double-clicked, the content of the selected file would be displayed in the right panel. Participants had to close the current file in order to open the other file since only one file could be displayed at a time. The search task was displayed one at a time in the bottom left panel. Participants entered the name of the apartment that they recommend for each client and clicked "enter" button to finish that task. Participants repeated the same procedure (with different search tasks) until they finished all eight tasks.

Luminance of icons was manipulated to reflect contextual information (recency and frequency) of icon access. We set five levels of luminance from the brightest to the darkest. In the very beginning of the experiment, luminance of all icons was identical (the darkest luminance level) for all conditions. In the L and L-SM conditions, however, luminance of icons started changing once the experiment began and participants started accessing icons. Base-level activation values in ACT-R cognitive architecture were calculated for each of the 20 icons, and were normalized and categorized to one of the five luminance levels every time a participant opened or closed a file. To participants' perspective, whenever an icon access was made, luminance level of each of all icons (including the one just accessed) changed (some became brighter while others became darker depending on when and how often it was accessed).

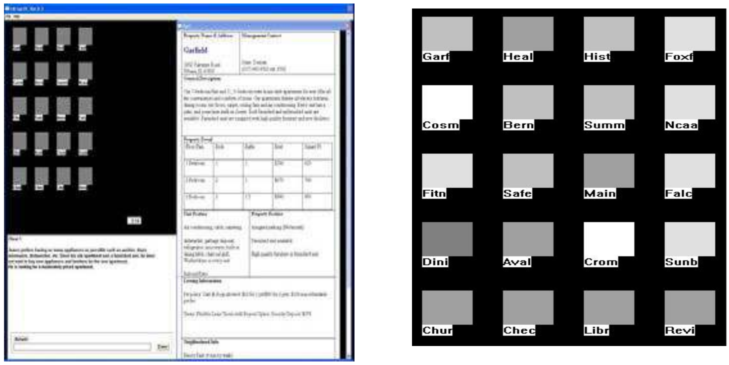


Fig. 2. Interface (left) and close-up view of 20 icons with luminance feature (right)

The use of recency and frequency as contextual information is relevant to the previous findings on human memory in rational approaches. In rational perspectives, sensitivity of human memory to frequency and recency values is adaptive since they predict the likelihood of encountering the same items in the future, which can be treated as a measure of the memory's future usefulness [5, 6]. In the subsymbolic level of ACT-R cognitive architecture, activation value (1) of a chunk (unit of declarative memory) reflects the degree to which past experience and current context indicate that chunk will be useful at any particular moment. Particularly, base-level activation (2) of chunk reflects the frequency (n) and recency (t) of its previous use. We focused on recency and frequency of information use (icon access) as predicting measure of future usefulness of information and provided them through luminance feature to support re-finding information.

$$A_i = B_i + \varepsilon \qquad \text{Bi: The base-level activation, } \varepsilon\text{: The noise value} \tag{1}$$

$$B_i = \ln\left(\sum_{j=1}^{n} t_j^{-d}\right) \tag{2}$$

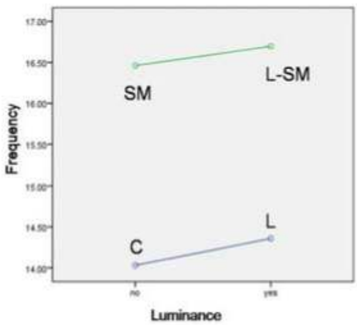
n: The number of presentations for chunk i.
tj: The time since the jth presentation.
d: The decay parameter

3 Results

A number of dependent measures were analyzed to measure performance as well as process. Accuracies were measured by assigning scores (from 1 to 4) depending on how close the answer was to the 'ideal' answer. Search times measured the total time to finish all eight tasks. We did not find significant difference between conditions in accuracies and average search times, which was probably due to a ceiling effect.

3.1 Access to Knowledge In-the-World

Despite the lack of significant difference in accuracies and search times, we found interesting differences in the search processes between conditions. Fig. 3 shows the



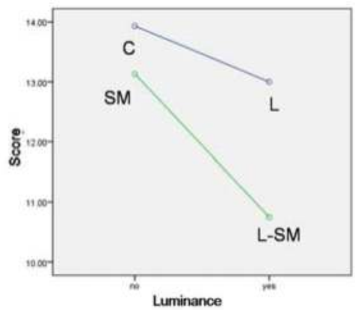


Fig. 3. Frequency of icon access (left) and result of icon-title-recall test (right)

average number of icon accesses per task. Participants with fixed-location icons (SM and L-SM condition) made significantly more accesses ($F(1, 60) = 4.641$, $p = .035$) than participants with randomized-location icons did. The effect of luminance was not significant.

The result was consistent with our prediction, which assumes that as the cost of accessing perfect knowledge in-the-world (KIW) increases, the cognitive system may satisfice with imperfect knowledge in-the-head (KIH). Randomizing locations (in the Control and Luminance conditions) increased external search costs, thus leading to fewer accesses to icons.

3.2 Access to Knowledge In-the-Head

Increased external search costs also implied that participants would rely more on internal memory. In order to test this, we conducted separate ANOVAs on the results from the memory tests of icon titles and locations across conditions. In the icon-title-recall test (Fig. 3), there was a significant main effect of luminance ($F(1, 58) = 4.560$, $p = .037$) as well as location ($F(1, 58) = 3.856$, $p = .054$). Participants with the luminance feature performed worse in recalling titles of the icons, suggesting that they relied less on titles during search. Similarly, participants with fixed-location icons performed worse than those with randomized-location icons. The interaction between luminance and location was not significant.

Results were again consistent with our predictions. Participants without the luminance feature performed better on the icon-title-recall test because remembering titles helped them to re-find the icons. When icon locations were randomly changed, they had to rely more on memory of the titles because the location cues were not available. On the other hand, the luminance feature has provided a useful cue for identification of target icons in addition to the titles. Participants with luminance features recalled fewer titles when location cues were not available (comparing participants in the C and L conditions) and when location cues were available (comparing participants in the SM and L-SM conditions). In either case, the presence of the luminance cues induced less reliance on the titles to re-find the icons.

The results of the icon-location-recall test provided further support to this claim. Participants in the SM condition recalled the locations significantly better than participants in the L-SM condition (M_{SM} = 22.75, $M_{L\text{-}SM}$ = 31.3125, p = .035). Lower scores mean better performance (the provided answer was either correct or closer to the correct location) in the test. Given that participants in the SM and L-SM conditions showed similar frequencies of icon accesses (Fig. 3), better recall of icon locations in the SM condition (no luminance) could be attributed to participants' higher reliance on spatial memory. Participants in the L-SM condition (with luminance), on the other hand, were worse in recall of icon locations, which could be attributed to their lower reliance on spatial memory. It is possible that the luminance feature, when combined with spatial memory, led to less precise encoding of spatial memory (i.e., remembering that the icon is somewhere at the top of the screen rather than remembering its exact location).

A notable point is that participants were not informed that they would be tested on either their memory of titles or locations of icons, and thus there was no reason to believe that they would have made any intentional effort to memorize them. Therefore, the scores in both tests reflected the strategic choice of utilizing knowledge-in-the-head vs. knowledge-in-the-world, as it emerged from the process of natural, dynamic interactions with the different interfaces.

3.3 Transition of Icon Access

In addition to the above mentioned findings, there were a couple of findings from the analysis of icon access transitions which was designed to identify patterns of search strategies. Transition data (Table 2) shows how a current icon access predicts a next icon access. For example, the number 52 in cell (1, 2) (in the first row and the second column in the table) means there were 52 times accesses to d2 (icon2) right after accesses to d1 (icon1) when access frequencies were aggregated for all participants in the SM condition. It also means that many participants accessed d1 located in the top left position in the grid, and then they accessed d2 which was located right next to d1.

A couple of interesting patterns were observed. First, in both the SM and L-SM conditions with fixed location icons, cells on the diagonal have relatively larger numbers (numbers marked bold), which suggests that participants did sequential search by accessing icons from top left toward bottom right (d1-d2-d3-…-d19-d20). This sequential search seems to be a strategy to minimize perceptual-motor efforts in search and to learn titles or locations of icons by accessing each one of them at a time. Sequential search might have been a strategy to easily identify icons already accessed and avoid redundant accesses.

This strategy might have been especially useful in the earlier sessions (task 1-4) when titles and locations were not fully encoded enough to be easily retrieved. However, as learning of titles and locations occurred over time, it is likely that participants were able to retrieve and use them for search in later sessions (task 5-8) of the experiment.

To further investigate the changes on the sequential search pattern over time, we split the table of the SM condition into two parts: part 1 (earlier sessions) and part 2 (later sessions). As expected, the sequential search pattern was more apparent

Table 2. Icon locations in the interface (left) and transition data of the SM condition (right)

d1	d2	d3	d4
d5	d6	d7	d8
d9	d10	d11	d12
d13	d14	d15	d16
d17	d18	d19	d20

	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10	d11	d12	d13	d14	d15	d16	d17	d18	d19	d20
d1	2	**52**	5	24	9	5	6	1	0	0	0	5	0	3	3	4	5	0	1	5
d2	15	0	**38**	19	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1
d3	7	8	1	**34**	5	1	2	1	0	0	1	0	0	1	1	1	0	0	0	1
d4	16	9	12	5	**61**	9	4	2	0	1	2	3	3	1	1	10	3	2	2	6
d5	10	2	1	30	1	**55**	6	1	4	1	1	6	1	5	1	12	7	1	2	4
d6	4	0	0	9	40	5	**55**	1	3	1	1	2	1	1	4	2	3	0	1	2
d7	10	0	0	4	6	44	6	**16**	5	5	9	27	0	17	3	6	2	1	0	2
d8	1	0	0	1	0	1	8	0	**15**	0	2	2	0	0	0	0	0	0	0	0
d9	1	0	1	0	1	1	11	5	0	**18**	6	2	2	2	0	2	0	0	0	1
d10	0	0	0	1	0	0	3	2	15	0	**21**	3	3	1	2	0	0	1	0	0
d11	1	0	0	1	3	0	9	0	5	20	0	**35**	1	2	1	1	0	0	0	2
d12	3	0	1	2	3	3	18	0	2	2	33	7	**25**	23	8	5	11	2	0	5
d13	2	1	0	0	0	0	1	0	0	1	1	17	0	**31**	1	1	3	1	0	0
d14	8	0	1	2	1	3	18	0	0	2	0	19	18	5	**56**	8	6	2	1	5
d15	6	0	2	2	3	4	0	0	1	0	1	5	2	37	4	**63**	6	1	0	3
d16	8	1	0	9	10	2	7	0	1	1	1	9	0	17	41	10	**53**	2	1	6
d17	11	0	1	3	4	4	5	0	1	0	1	3	2	6	10	34	2	**36**	4	4
d18	4	0	0	2	0	0	0	0	1	0	0	0	0	0	1	7	21	0	**21**	9
d19	2	0	0	1	2	0	1	0	0	0	0	2	0	0	1	3	2	11	0	**18**
d20	12	1	1	4	5	2	3	1	0	0	1	6	1	3	3	11	6	6	10	1

in part 1 (40.49%) than in part 2 (24.58%). Although the sequential search pattern still appeared in part 2 possibly due to the fact that some of the ‘more relevant’ files were next to each other and due to this proximity, icon accesses might have been still sequential even when participants were using other strategies. The similar pattern was observed for the L-SM condition.

We were also interested in icon accesses made to the more relevant files versus the less relevant files over time to verify that participants were using the features of icons effectively in search. Patterns of transition data seem to support this idea. For instance, in the C condition with random-location icons, icon accesses were suboptimal in the earlier sessions. Only 46.42% of the total icon accesses were made to the files which contained information more directly relevant to correct answers. However, in the later sessions, 63.51% (17% increases) of the total icon accesses were made to those more relevant files. This pattern suggests that participants optimized icon accesses over time by utilizing features of icons in the absence of location cue. Although it does not directly show exactly what features they were using for search, the pattern of focused accesses is clear enough to support the optimization of information search over the sessions.

In addition, as similar as in the SM condition, participants initially made more frequent accesses to the icons in the top left area (when the same data was coded based on location) in the earlier sessions (19.70%). It seems that participants were learning the titles while minimizing perceptual-motor efforts (by accessing particular locations redundantly) until the titles of icons were encoded enough to be retrieved. This pattern was weakened in the later sessions (9.33%), suggesting that they started using titles of icons more actively over time.

4 Discussions

The results from the experiment were consistent with the theory of soft constraints, which was derived based on the perspective that human information processing is highly adaptive to the characteristics of the external environment. Randomized locations had increased the cost of accessing knowledge in-the-world, and thus induced a shift to more accesses to knowledge in-the-head (more memory encoding and retrieval of icon titles and locations). Fixed locations, on the other hand, had reduced cost of accessing knowledge in-the-world, and thus induced a shift to fewer accesses to knowledge in-the-head. Similarly, the luminance feature had reduced the cost of accessing knowledge in-the-world, and resulted in less reliance on knowledge in-the-head. Results therefore supported the notion that presentation of contextual information not only led to different external representations of information, but also led to changes in the processes of the users.

The fact that participants in the L-SM condition, with significantly less episodic and spatial memory encoding, could attain comparable search performance is worth noting. The luminance feature apparently guided re-finding of icons by highlighting and providing contextual information. Results highlight the importance of integrating bottom-up visual features and top-down contextual information to augment knowledge in-the-head. Participants without luminance and/or location cues compensated for the poor interface support by using more of their memory of titles and location. It was therefore possible that the luminance feature could reduce the memory load, but this hypothesis apparently needs to be further tested.

One implication of this study is that representations and processes are interdependent and are both essential for interface design. Given different interfaces, participants could adapt their strategies (processes) to attain the same level of performance. It is therefore superficial to argue that a particular interface is a good representation without carefully examining also the processes that interact with that representation. The study also shows that systematic testing and understanding of how people utilize contextual information (e.g. recency, frequency, types of contents) is critical in deciding what are their best representations (e.g. luminance, colors, etc), and when these representations are best in what situations (e.g. short-term, long-term, intermittent, persistent use).

References

1. Alvarado, C., Teevan, J., Ackerman, M.S., Karger, D.: Surviving the Information Explosion: How People Find Their Electronic Information. In: AI Memo 2003-006, Massachusetts Institute of Technology Artificial Intelligence Laboratory (2003)
2. Anderson, J.R.: The Adaptive Character of Thought. Erlbaum, Hillsdale (1990)
3. Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An Integrated Theory of the Mind. Psychological Review 111, 1036–1060 (2004)
4. Anderson, J.R., Lebiere, C.: The Atomic Components of Thought. LEA, Hillsdale (1998)
5. Anderson, J.R., Milson, R.: Human Memory: An Adaptive Perspective. Psychological Review 96, 703–719 (1989)
6. Anderson, J.R., Schooler, L.J.: Reflections of the Environment in Memory. Psychological Science 2, 396–408 (1991)
7. Capra, R., Pérez-Quiñones, M.A.: Re-Finding Found Things: An Exploratory Study of How Users Re-Find Information. Technical Report cs.HC/0310011, Computing Research Repository (2003)
8. Ehret, B.: Learning Where to Look: Location Learning in Graphical User Interfaces. In: Proc.of CHI'2002 Conference on Human Factors in Computing Systems, April 20–25, pp. 211–218 (2002)

9. Fleetwood, M.D.: Computational Modeling of Icon Search. Unpublished master's thesis, Rice University, Houston, Texas (2001)
10. Fu, W.-T., Gray, W.D.: Memory versus Perceptual-Motor Tradeoffs in a Blocks World Task. In: Proceedings of the 22nd Annual Conference of the Cognitive Science Society. Erlbaum, Mahwah (2000)
11. Fu, W.-T., Gray, W.D.: Suboptimal Tradeoffs in Information-Seeking. Cognitive Psychology 52, 195–242 (2006)
12. Fu, W.-T., Pirolli, P.: SNIF-ACT: A cognitive model of user navigation on the World Wide Web. Human-Computer Interaction 22(4), 355–412 (2007)
13. Gray, W.D., Fu, W.-t.: Ignoring Perfect Knowledge In-the-World for Imperfect Knowledge In-the-Head: Implications of Rational Analysis for Interface Design. CHI Letters 3(1), 112–119 (2001)
14. Gray, W.D., Sims, C.R., Fu, W.-T., Schoelles, M.J.: The Soft Constraints Hypothesis: A Rational Analysis Approach to Resource Allocation for Interactive Behavior. Psychological Review 113(3), 461–482 (2006)
15. Houde, S., Salomon, G.: Working towards Rich and Flexible File Representations. In: Proceedings of the CHI 1994 conference companion on Human factors in computing systems, pp. 9–10 (1993)
16. Ishii, H., Ullmer, B.: Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. In: Proc.of Conference on Human Factors in Computing Systems (CHI 1997), March 1997, pp. 234–241. ACM, Atlanta (1997)
17. Jones, W., Dumais, S.: The Spatial Metaphor for User Interfaces: Experimental Tests of Reference by Location versus Name. ACM Transactions on Office Information Systems 4(1), 42–63 (1986)
18. Lansdale, M.: The Psychology of Personal Information Management. Applied Ergonomics 19(1), 458–465 (1988)
19. Mandler, J.M., Seegmiller, O., Day, J.: On the Coding of Spatial Information. Memory & Cognition 5, 10–16 (1977)

Adaptive Work-Centered and Human-Aware Support Agents for Augmented Cognition in Tactical Environments

Martijn Neef[1], Peter-Paul van Maanen[1,2], Peter Petiet[1], and Maartje Spoelstra[1]

[1] TNO Defence, Security and Safety, The Netherlands
[2] Vrije Universiteit, Amsterdam, The Netherlands
{martijn.neef,peter-paul.vanmaanen,peter.petiet,
maartje.spoelstra}@tno.nl

Abstract. We introduce a support system concept that offers both work-centered and human-aware support for operators in tactical command and control environments. The support system augments the cognitive capabilities of the operator by offering instant, personalized task and work support. The operator obtains support by entering into a collaborative agreement with support agents. Such an agreement creates opportunities to establish adaptive capabilities and human-aware support features. We describe the concept in and propose an experimental design to evaluate its effectiveness in tactical environments.

1 Introduction

Tactical command and control environments present demanding challenges to operators because of their task complexity, time-criticality and the risk associated with poor decisions. With the increasing complexity of defense missions and the ongoing pressure to reduce manning in defense environments, operators will have more difficulty in handling their tasks than before. Operators will continue to face new challenges in properly focusing attention, keeping correct situation awareness and preventing overload situations. Conventional technology-driven support systems are not capable and adaptive enough to fulfill such support demands. There is a need for novel support concepts that focus on augmenting the operator's cognitive capabilities.

Support systems for tactical environments come in many flavors, but many recent developments show support systems that resemble active companions rather than conventional decision aids. The key feature of such companions is adaptivity. Adaptive capabilities for support systems come in two varieties: (a) adaptivity based on the behavior and performance of the human operator, and (b) adaptivity based on the state of the environment. The first variety could be called *human-aware* adaptivity, because it requires the support system to have a certain degree of awareness about the behavior and motivations of the human operator. The second variety could be called *work-centered* adaptivity, because it revolves about events in the environment, and thus mainly requires knowledge about the work domain. Both types of adaptivity result in changes in the interaction between the human operator and his support systems. This includes every option from a change in notification method to a change in the division

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 68–77, 2009.

of labor between operator and support system. We believe that a support system should include both types of adaptivity, and be readily configurable, so that the operator can benefit from adaptive capabilities at the right moment, and in the right form.

We propose an support system based on work on human aware agents (Van Maanen, 2008), adaptive autonomy (Van de Vecht, 2007), and work-centered support systems (Scott, 2005), that specifically targets the attention allocation problem and task overload issue of tactical officers. The concept gives the officer options to arrange and configure support capabilities at will in the form of autonomous software agents that act as versatile companions. It is the responsibility of the human user to instantiate support agents, and to tune them to the task at hand. By making it the responsibility of the operator to create his own support functions, we lessen the possibility of ‘automation surprises’ (Sarter, 1997), and make the operator better aware of his system surroundings, goals, and states.

1.1 Work-Centered Support Systems

Work-centered support system design uses methods from cognitive system engineering (cognitive work and task analysis, work-centered design) to understand and support operators in complex work environments. It adds agent technology to create practical support tools that help to reduce work complexity and keep the user in control of his work (Scott, 2005). Traditional user interface design is primarily focused on the capabilities of the system, and forms the interface so that system features are as accessible as possible. Work-centered support system design seeks to shape the working environment in such a way that it agrees with the user’s work, and make relationships between elements, constraints and affordances in the environment easier to observe. Support systems built on this premise are context sensitive, easily adaptable and tuned to the antics and ontologies of the human way of working.

A good example of a work-centered support system is the work by Eggleston on support agents for weather forecasters in a military airlift organization (Scott, 2002). Their design allows operators to create agents instantly for various types of work support. For example, operators can use visual signs to instruct watch agents to monitor certain events on the interface, such as the formation of thunderstorms. Other types of agents in the system monitor external information sources and manage the presentation of information to the user. By letting the operator self-initiate support agents and providing a high level of observability and directability of automation behavior, there is less chance of automation surprises, which makes the system easier to accept.

1.2 Human-Aware Support Systems

Van Maanen (2008) introduces a generic human attention-based task allocation (HABTA) agent component for tactical command and control environments. When the complexity of the in- and external environment increases, so will the information volumes for navigation, system monitoring, and tactical tasks (Grootjen, 2006). With expanding information volumes, keeping a proper attention focus becomes a challenging task, especially for novice operators and in situations where there is a need for frequent attention switches between tasks or objects.

Human-aware support agents augment the meta-cognitive capabilities of human operators. A HABTA-component is based on two cognitive models: one that describes the current allocation of a human's attention and one that prescribes the way his attention should be allocated. If there is a discrepancy between the outputs of both models, HABTA reallocates the tasks between the human and the agent. The HABTA-based agent presented in (Van Maanen, 2008) is able to monitor the complete environment, and when the agent decides that human attention is needed within a specific area, it draws the attention of the human operator. In other words, it has a high level of human-aware adaptivity (directs human attention), but a low level of work-centered adaptivity (does not perform the task itself).

1.3 Adaptive Autonomy

Adaptivity is an often-sought feature of support systems. Most recent work in this field use autonomous agents to model adaptive capabilities. Autonomy is a defining characteristic of intelligent agents (Jennings, 2000) and it usually interpreted as the amount of freedom an agent has from intervention by other agents, including humans (Barber, 2001). An agent heavily influenced in its decision making by other agents displays obedient behavior. An agent that does not allow any external influence in its decision making is maximally independent. By altering the amount of external influences on its decision making, an agent can self-adapt its autonomy, and effectively display adaptive autonomy. In this fashion, agents can actively select the level of autonomy that best fits the circumstances and their own objectives. Van der Vecht (2007) developed a model that implements this perception of autonomy by means of reasoning rules that decide on adoption or rejection of certain influences.

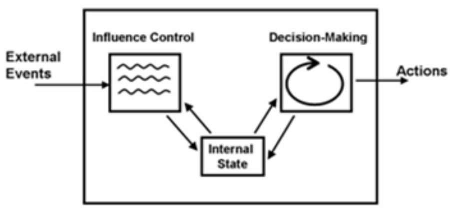


Fig. 1. Illustration how influence control can affect agent behaviour (Van der Vecht, 2007)

By giving the agent self-control over these rules, the agent can filter out certain events and communications and consequently adjust his own level of autonomy.

The approach of Van der Vecht to adaptive autonomy offers interesting perspectives for use in support system design, because it provides a comprehensive way to represent the rules that produce adaptive behavior. It also returns the responsibility for adaptive behavior itself to the agent, while giving the operator straightforward means to convey preferences (i.e. by manipulating the influence control settings of the support agent so that the boundaries of the autonomy of the agent become clear). Van der Vecht uses social contracts, as proposed in the OperA framework (Dignum, 2004), to convey organizational knowledge into the influence control rules. Upon entering an organization, the agent signs a contract that specifies its role and hierarchical relations to other actors. Similarly, we propose to use interaction contracts between operator and support agent to specify adaptive behaviors in a support setting.

2 Combining Work-Centered and Human-Aware Adaptivity

We combine elements from work-centered design, human-aware support systems and adaptive autonomy to form a novel approach to tactical C2 support system design. Our basic stance is that the operator should be responsible for creating his own set of support agents. The operator creates agents by means of tasking descriptions and interaction contracts. These instructions give the agents a basis for their behavior. Because the operator actively creates an agent for a specific task, and explicitly communicates his expectations about the agent's behavior, it will be easier for the operator to remain in control over his suite of support tools.

There are two types of agents that the operator can create, namely work-centered agents and human-aware agents. Work-centered agents deal with events in the work domain, such as the monitoring of a certain geographical map for dangerous events. Human-aware agents deal with the meta-cognitive aspects of the operator, such as workload management and attention strategies. Most human-aware tasks will require a normative model for assessing the operators' state or actions. Therefore, the use of human-aware task agents will often give rise to the use of work-centered agents. Jointly, human-aware task agents and work-centered task agents make it possible to create comparative support features, such as human attention assessment. Figure 2 shows the interactions between user and agents.

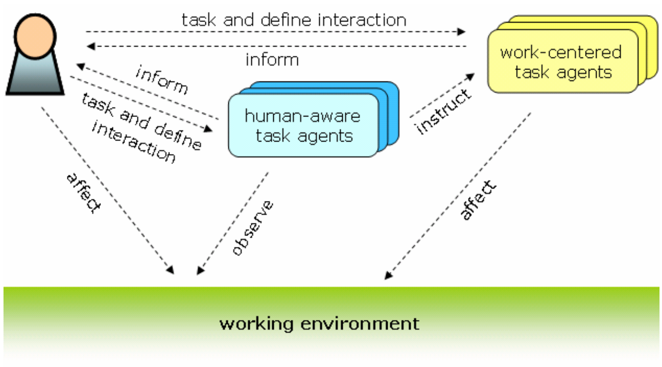


Fig. 2. Interactions between the operator, human-aware agents and work-centered agents

2.1 Creating Support Agents

The creation of support agents consists of three steps for the operator: *defining the support task*, *defining the interaction,* and *instantiating the agent* (see Figure 3). In the first step, the operator expresses a functional task description for the support agent. This description includes essential aspects such as objectives, start- and stop-criteria, and pre- and post-conditions. How the operator actually communicates this tasking description is of less importance, as long as the process uses visualizations and terms from the work-domain itself. Tactical C2 officers usually have a map-based work environment, and most of their tasks have geographical attributes ('monitor this area', 'follow this track'). It would make sense to use a tasking method that uses objects and terms from that environment, which seamlessly integrates with the operators' own way of working. For example, for a work-centered agent, the operator could mark out an area using his input device, and make a selection from a predefined set of typical support tasks (e.g. 'monitor this area for suspect tracks', see Table 1). This would be a straightforward manner for the operator to request help, without having to leave his work domain to enter a 'programming process'. On the side of the agent, the visual tasking would be translated in a set of actions and pre- and post-conditions.

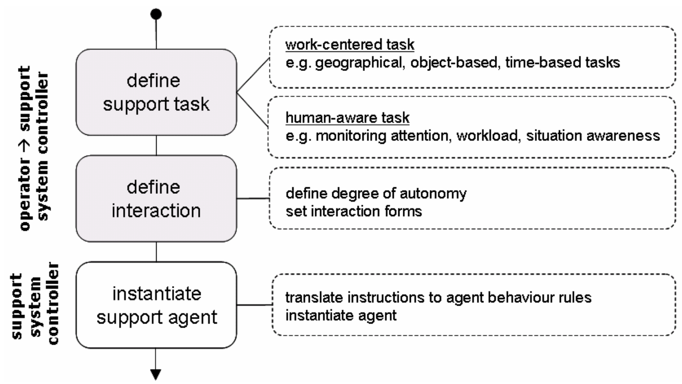


Fig. 3. Definition and instantiation of support agents

The second part of the agent creation process deals with defining the interaction between operator and support agent by means of an interaction contract. The interaction contract defines the desired behavior of the agent towards the operator. It delimits the extent of its autonomy (in how far the agent can take control of its own behavior), it defines communication requirements towards the operator (what should the agent communicate and under which conditions), and sets collaborative aspects (what is the division of labor between agent and operator). It defines the degree of autonomy that the agent is granted, and states in how far, and under which conditions the agents may perform its task independently from the operator. The interaction contract describes per task, when it is permitted. For example, if the agent has been instructed to monitor an area and act upon certain events, then the interaction contract may specify whether the operator needs to be consulted before acting. In this case the contract can contain conditions related to the agent's awareness of the user's task execution of current attention division and based on these conditions determines either to interrupt the user, wait until a more appropriate moment, or execute his task anyway. In the tasking description, this actually sets the pre-condition.

The support agents use a reasoning model to decide on their behavior. In this model (Van der Vecht, 2007) there is a distinction between 'influence control' and 'decision making' (see figure 2). The autonomy of an agent is determined by the

Table 1. Examples of tasking and interactions descriptions in pseudo code

Tasking description (work-centered task)		
task	Monitor (area) for contacts	
type	work-centered	
agent	WCA_1	
subtasks	MONITOR (area) CLASSIFY (object) REACT (object) INFORM (actor, form, status)	observe area classify object activate countermeasures provide information
rules	MONITOR (area) for contacts IF status(contact) = new, THEN CLASSIFY (contact) IF status(contact) = suspect THEN REACT (contact)	
preconditions	main: IF operator permission THEN do (main_task) for all subtasks: interpret interaction contract determine permission and information requirement	

Tasking description (human-aware task)		
task	Attention support (area)	
type	human-aware	
agent	HAA_1	
subtasks	OBSERVE (actor) COMPARE (actor1, actor2) ASSIGN (objects, actors) INFORM (actor, form, status)	observe actions of actor compare actions of actors allocate actions to actor provide information
rules	O1 = OBSERVE (operator) O2 = OBSERVE (WCA_1) IF difference (O1, O2) > threshold THEN INFORM (operator, form, status) ASSIGN (contacts, WCA_1)	
preconditions	main: IF operator permission THEN do (main_task) for all subtasks: interpret interaction contract determine conditions and information requirements	

Interaction description	
parties	operator, WCA_1
task	Monitor (area) for contacts

rules	relevant trigger: workload
	MONITOR: allowed(agent, MONITOR)
	CLASSIFY: allowed(agent, CLASSIFY)
	IF status(contact) = suspect THEN
	required(agent, INFORM (suspect), warning)
	REACT: conditional
	IF workload (operator) > threshold2 THEN
	allowed(agent, REACT)
	required(agent, INFORM (operator, notification, action)
	IF workload (operator) > threshold1 THEN
	required(agent, INFORM (operator, warning, workload)

amount of influence that other actors have on the agent's decision making. The influence-control function actively regulates which events and commands reach the decision-making process. It filters out those external events that should not influence the decision making, and thus it indirectly controls the agent's behavior. Similarly, the influence-control function can adapt the agent's behavior by altering its settings. For example, if the agent should take the workload of the operator into account in its decision making process (human-awareness), the influence control rules can be adjusted so that cues pertaining to the state of the operator are admitted to the agent.

Table 2 show examples of rules that govern the influence control. If the former part of the rule (the head) of the rule agrees with the condition (the guard), then the latter part (the body) is performed. Adding a 'goal' boils down to obliging the agent to

Table 2. Three examples of reasoning rules for setting influence control

'Whenever status-suspect then agent is obliged to do inform-operator-about-status'
observation(status-suspect) → TRUE | AddGoal(send(operator, inform, new-suspect-contact))
'Whenever workload-high then agent is permitted to do propose-operator-contact-takeover'
observation(workload-high) → TRUE | AddBelief(permitted(propose-operator-contact-takeover))
'Whenever takeover-request from operator then agent is obliged to do accept-request'
message(operator, request, raise-support-level) → TRUE | AddGoal(takeover(Task))

perform a certain action. Adding a 'belief' gives the agent a notion of permissible actions. By giving the agents simple rules like these, the operator can communicate his preferences, and easily change agent behavior.

3 Experimental Validation Strategy

In this section, we describe an experimental design for a first-order validation of the concept. We draw inspiration from earlier work on active support systems such as the TANDEM experiments in the TADMUS project (Sycara, 2004), and the aforementioned studies on HABTA-based agents (Van Maanen, 2008), and propose an experimental setting in which we evaluate the operator performance under different saturation conditions. When stressed enough, the attention of the operator will become more and more saturated and will need support to achieve his objectives.

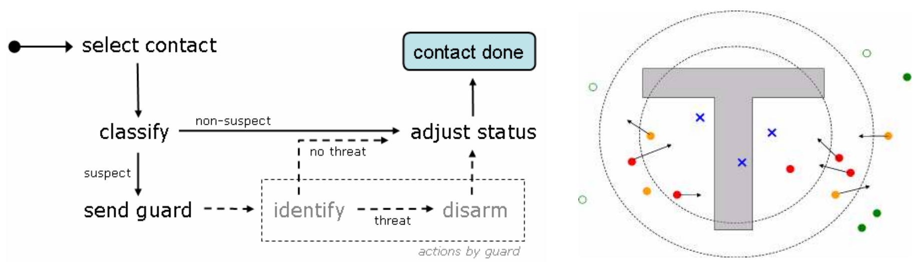


Fig. 4. Left: Task sequence in the experiment. Right: Basic scenario environment layout

We propose a common tactical scenario: the protection of a High Value Target (HVT). The HVT is an airport, threatened by hostile entities. The airport is defended by mobile guards, who can intercept individuals that pose a security threat. A commander with global overview of the HVT area instructs the guards. Sensor networks around the HVT reveal movements of individuals, but cannot reveal whether their intent is malicious. If the commander has reasons to believe that a certain individual poses a threat to the airport, then he can instruct guards to identify, and, if necessary, disarm that individual. The challenge for the operator is to keep track of all contacts on the screen, while commanding his guards in such a way that they do not miss out on any threats to the airport. This implies two vital decisions on the part of the commander: (a) deciding whether a contact is suspect or not, and (b) deciding whether guards should intercept the contact. Figure 4 shows the sequence of tasks.

The operator will have a limited number of options to create support agents. Work-centered agents can do the exact same tasks as the operator (*observe, select contact, classify, send guard, adjust status*), but the operator decides which tasks the agent may perform. The operator can opt to have a work-centered support agent to help monitor a certain area (*observe* and *classify)*, or assist by also actually instructing guards (*observe, classify* and *send guard).* This selection equals setting the degree of autonomy. In addition, the operator can include a limited number of human-aware features, such as *attention monitoring (*based on comparing classification assessments between operator and a work-centered agent) and *workload management* (based on the time it takes for a contact to be classified by the operator). Furthermore, the tasking interface will give the operator a number of ways to divide the labor between himself and the agents. The most reasonable options would be via spatial arrangements (sectors, perimeters or free-form areas that the operator can draw on the screen), based on perceived threat (the operator deals with all suspect contacts, the agent is responsible for monitoring unknown or safe contacts).

By varying three scenario parameters, we can challenge the operator, and evaluate the behavior of the support agents. By increasing the number of contacts on the screen (*volume)*, or increasing the movement speed of contacts *(time)*, we can drive the operator into a work overload situation. We can also make the classification task harder (*complexity*) by letting contacts follow paths that are more diverse. A contact that takes a long detour along the area before turning towards the airport will be harder to assess than a contact that follows a straight line towards the airport. Another option to increase complexity would be to include more non-threatening entities. Varying settings of these parameters will likely lead to a different use of the support system, and thus give an interesting insight in the value of such a support system in tactical environments. For instance, a higher level of complexity could possibly lead to more reliance on human-aware support, while the operator will most likely benefit more from work-centered agents under higher volume circumstances. Exploring the range of use scenarios is essential given the many different types of missions in tactical command and control.

We expect that this experimental setting will reveal critical usability and efficiency indicators, and give us a better understanding of the relationship between scenario-type and most favorable type of support, in terms of adaptive capabilities and user interaction.

4 Summary and Conclusion

In this paper, we have introduced the approach of combining work-centered and human-aware adaptivity in support agents for operators working in tactical command and control environments. We can summarize the approach in the following set of design requirements:

- Support functions are embodied by autonomous software agents.
- It should be possible for the operator to create a support agent for a specific task at his will. This includes work-centered tasks (tasks that relate to the domain), and human-aware tasks (tasks that relate to the behavior of the operator), or the combination thereof.

- It should be possible for the operator to observe and configure the behavior of the support agents, including the limits of its autonomy and specific interaction requirements via suitable controls.
- The support agent should have a suitable normative model of user behavior and (part of) the work domain. These models may be available by design, or given to the support agent by the operator through simple instructions (for instance, performance indicators, geographical cues).
- Human-aware agents must be able to instruct existing work-centered agents, or create such agents themselves. This allows a human-aware agent to autonomously reallocate tasks from the human operator user to the support suite in the case of suboptimal performance by the operator.
- Work-centered agents instantiated by a human-aware agent should behave according to the interaction contract set between the operator and the human-aware agent.

We expect that this strategy is going to be very effective in domains where there is limited time and resources for the execution of multiple tasks, but where the operating environment provides high volumes of information.

This approach is a first step towards a framework that can cater both technical and operational demands, and that is open enough to fit all sorts of cognitive requirements. We believe that our approach is novel and worthwhile, because it joins several established lines of thought on support system design. It augments cognition by providing assistance in meta-cognitive processes (human-aware) and lessening decision making complexity by offering work-centered task support. By keeping a distinction between human-aware tasks and work-centered tasks, it becomes easier to design tailored, adaptive support. Clear interaction contracts between operator and support agents define the boundaries of agent behavior, and minimize the chances on automation surprises. The practice of interaction contracts agrees with most cognitive system engineering proposition, and help to maintain observability and directability of system functions (Klein, 2004). By giving the operator agent controls that use visualizations and ontologies from his work-domain, user acceptance will be easier to achieve.

The approach needs practical evaluation, and this will most likely reveal many implications. We may inadvertently introduce many new issues. By effectively shifting the agent design process from designer to user, we lessen many design challenges that stem from user modeling. At the same time, this implies that the user becomes partly responsible for system management as well. The management of a large number of autonomous support agents will require extra attention and take up valuable work capacity. The use of contracts to instantly create support agents is a fascinating concept, but it will require an extensive exploration into user tasking controls, interaction formats and management procedures.

Acknowledgements

The research reported here is funded by the Dutch Ministry of Economic Affairs (Interactive Collaborative Information Systems (ICIS) project, grant nr. BSIK03024) and by the Dutch Ministry of Defense (Programme number V524).

References

1. Barber, K.S., Martin, C.E.: Dynamic adaptive autonomy in multi-agent systems: Representation and justification. International Journal of Pattern Recognition and Artificial Intelligence 15(3), 405–433 (2001)
2. Grootjen, M., Neerincx, M.A., van Weert, J.C.M.: Task-based interpretation of operator state information for adaptive support. In: Foundations of Augmented Cognition, 2nd edn., Strategic Analysis, Arlington, Virginia, pp. 236–242 (2006)
3. Jennings, N.R.: On agent-based software engineering. Artificial Intelligence 117(2), 277–296 (2000)
4. Klein, G., Woods, D.D., Bradshaw, J.M., Hoffman, R.R., Feltovich, P.J.: Ten Challenges for Making Automation a ”Team Player” in Joint Human-Agent Activity. IEEE Intelligent Systems 19(6), 91–95 (2004)
5. van Maanen, P.-P., de Koning, L., van Dongen, K.: Design and Validation of HABTA: Human Attention-Based Task Allocator. In: Mühlhäuser, M., Ferscha, A., Aitenbichler, E. (eds.) Proceedings of the First International Workshop on Human Aspects in Ambient Intelligence, Constructing Ambient Intelligence: AmI 2007 Workshops Proceedings. Communications in Computer and Information Science (CCIS), vol. 11, pp. 286–300. Springer, Heidelberg (2008)
6. Sarter, N.B., Woods, D.D., Billings, C.E.: Automation Surprises. In: Salvendy, G. (ed.) Handbook of Human Factors and Ergonomics, 2nd edn., pp. 1926–1943. Wiley, New York (1997)
7. Scott, R., Deutsch, S., Kazmierczak, T., Kuper, S., Roth, E., Malchiodi, E., Eggleston, R., Whitaker, R.: Using Software Agents in a Work Centered Support System for Weather Forecasting and Monitoring. In: Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting, Baltimore MD. Human Factors and Ergonomics Society, Santa Monica (2002)
8. Scott, R., Roth, E.M., Deutsch, S.E., Malchiodi, E., Kazmierczak, T., Eggleston, R., Kuper, S.R., Whitaker, R.: Work-Centered Support Systems: A Human-Centered Approach to Intelligent System Design. IEEE Intelligent Systems 20(2), 73–81 (2005)
9. Sycara, K., Lewis, M.: Integrating intelligent agents into human teams. In: Salas, E., Fiore, S. (eds.) Team Cognition: Understanding the Factors that Drive Process and Performance, pp. 203–232. American Psychological Association, Washington (2004)
10. van der Vecht, B., Dignum, F., Meyer, J.-J.C., Neef, M.: A Dynamic Coordination Mechanism Using Adjustable Autonomy. In: Sichman, J.S., Padget, J., Ossowski, S., Noriega, P. (eds.) COIN 2007. LNCS, vol. 4870, pp. 83–96. Springer, Heidelberg (2008)

Designing Cognition-Centric Smart Room Predicting Inhabitant Activities

A.L. Ronzhin, A.A. Karpov, and I.S. Kipyatkova

St. Petersburg Institute for Informatics and Automation, 39, 14th line, St. Petersburg, Russia
{ronzhin,karpov,kipyatkova}@iias.spb.su

Abstract. Assignment of easy-to-use and well-timed services staying invisible for a user is one of important features of ambient intelligent. Multimodal user interface capable to perceive speech, movements, poses and gestures of participants in order to determinate their needs provides the natural and intuitively understandable way of interaction with the developed intelligent meeting room. Awareness of the room about spatial position of the participants, their current activities, roles in a current event, their preferences helps to predict more accurately the intentions and needs of participants. Technological framework, equipment and description of technologies applied to the intelligent meeting room are presented. Some scenarios and data structures used for a formalization of context and behavior information from practical human-human, human-machine and machine-machine interaction are discussed.

Keywords: ambient intelligence, cognitive-centric design, multimodal interfaces, context awareness, smart home, intelligent meeting room.

1 Introduction

An idea of recognition of a current situation and behavior of a user, as well as an unobtrusive satisfaction of his needs underlies the “ambient intelligence” (AmI). These tasks deal with three directions of science and technique: ubiquitous computing, ubiquitous communication, and multimodal interfaces [1]. Integration of diverse computation, information and communication resources into a united framework is one of the important issues at design of ambient intelligence and it identifies the modern tendency to transition from smart devices to an ambient intelligent space. Multimodal interfaces provide natural and intuitively comprehensible interaction between a user and intellectual devices, which are embedded into the environment. All the means should be hidden, thus the user can see only the results of intellectual devices activities and concentrate attention on her/his work.

The so-called “smart home” (SH) is the most profoundly studied area from the domain of the ambient intelligence. A smart home can be defined as a dwelling house equipped with computational and informational technologies guessing and reacting upon needs of the inhabitants, working on comfort maintenance, security and good entertainment by means of control of domestic equipment and communications [2]. In spite of the fact that SH idea was well established in the late 1990s, SH technologies still have no popularity among potential users. It is Gann’s opinion [3] that the

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 78–87, 2009.

principle barriers to uptake are the following: high cost, difficulties of integration with contemporary domestic appliances and lack of usability.

Industry of SH should take into account a number of criteria to motivate customers in its advantages [3]. First, it is necessary to consider real needs of users. Secondly, development should be conducted at three levels at once: (1) generic technologies necessary for a base configuration applied to all appendices; (2) context-dependent technologies which can be adapted for the majority of homes; (3) personified technologies developed with respect to requirements of concrete customers and facilities. Thirdly, the offered decisions should be flexible, simple of use, reasonable of price, reliable, ease of installation and reconfiguration.

Taking into account functionality available to a user it is necessary to distinguish homes with smart devices and homes providing interactive computing both inside and beyond the home. Five classes of SH were proposed in [3]: (1) home, which contains stand-alone intelligent objects, actions of which are mutually independent; (2) home, which contains network of intelligent objects worked by own rules, but due to exchange information between one another, their functions are enhanced; (3) connected home has internal networks connected to external ones and allows interactive and remote control of home systems, as well as access to services both inside and outside the house; (4) learning home, able to record and store data about activities patterns in order to use them in future; (5) attentive home, constantly recording data of family members localization in order to predict users' behavior, satisfying their needs and unobtrusively interact with them.

The classification takes into account not only functional abilities of SH, but also highlights different levels of communication within and beyond the house, starting from the simplest systems with mechanical toggle switches, and to complicated systems, forecasting user's behavior and interacting with him.

Ambient intelligence paradigm and the subsequent development of ambient intelligent space have led to formation of the so-called cognitive approach become of great importance for designing SH systems. Intelligent agents embedded into the environment, allow AmI distinguishing people behavior and reacting to their needs in an unobtrusive and even invisible way owing to intuitive interfaces [1]. The cognitive approach to design, with the aim of AmI, gives to the developer a way to compare demands and cognitive processes of the end user. This goal can be reached owing to intellectual agents networks, which provide synchronization of information streams, remote control and user notification, in the aggregate with adaptive interfaces and the means for achievement of full situational awareness [4]. The knowledge of a context helps to explain behavior of the user in a concrete situation and to train the system to react adequately.

Cognitive and perceptive systems of a human being automatically use a context for identification in everyday life. The context can be defined as a set of relevant conditions and ambient influences which make a certain situation unique and distinguished. The context implies a notion of weight of influencing factors, considered subconsciously by people but being beyond perception of artificial systems. The optimal approach to formalization of contextual information is an iterative two-step procedure: (1) collection of the detailed expert description of situations; (2) check of accuracy of automatic recognition of the certain contexts on the basis of previously created descriptions [5].

One of the basic directions within the AmI paradigm is a development of multimodal user interfaces providing natural and intuitively clear way of human-computer interaction. Speech, gestures, handwriting and other means, which are natural of inter-human dialogue, are applied to management of machines in multimodal interfaces [6]. Audio-visual information processing allows automatic identifying the person, his intentions, speech, movements and the current position. Due to the analysis of single modalities and their subsequent fusion on the semantic level it is possible to improve naturalness and robustness of interactions.

Multimodal integration combines semantic hypotheses coming from modules of single modalities processing and carries on joint interpretation of a user's command. For recognition of user's intention (communicative or practical) it is necessary to use information from different sources, and to consider various types of the contextual information. Three types of contexts are proposed to use during the fusion of multimodal information [7]: (1) context of a subject domain, which contains some a priori knowledge, the user's preferences, models of situations, descriptions of objects and subjects, their possible activities and their location to each other; (2) conversational context describing possible dialogues with the system and a current condition of a discourse; (3) visual context including the analysis of a direction of vision, gestures, actions of the user in the course of the observable situation. Data and knowledge, accumulated by the system in advance, determine an interpretation of user's intentions and the current situation. The result of recognition of a command is, theoretically, a multimodal representation of user's behavior, conclusion about his intentions and a succession of activities which are to be performed by the system.

There are more fields for applying areas of AmI besides that of a SH. One may state that requirements for a SH were worked out already, but requirements for meeting and lecture rooms are somewhat insufficiently standardized. Sometimes even users have no idea what they should expect from such intelligent rooms. The user cannot quickly become familiar with a simultaneous use of several wide-screen displays, interactive multimedia support, integration of mobile devices and others new resources. All these means and tools must be intuitively comprehensible. Usually such rooms function in a semi-automatic mode, and all the embedded systems are maintained by experts.

Peculiar features of multimodal interfaces exploiting and human behavior in the intelligent meeting room of SPIIRAS are regarded in the present paper. Testing of audio-visual processing is made during laboratory meetings and lectures. Besides, control of the equipment was realized on a basis of mobile-phone, while collecting of data concerning SPIIRAS members was performed with the aim of a mobile robot with a multimodal interface, moving along a corridor of the institute.

The next section presents basic features of the equipment and description of technologies applied to the intelligent meeting room. Apart from that, some scenarios and data structures used for a formalization of context and user behavior in the meeting room will be presented.

2 Technological Framework of Intelligent Meeting Room

A premises of 72 square meters located inside the institute building was supplied for intelligent meeting room. The room scheme and arrangement of base equipments are presented in Figure 1. Monitoring of the room is performed by 16 video cameras mounted on the walls, ceiling and tables and provides tracking of moving objects, face detection and other functions. Three T-shape 4 channel microphone arrays mounted on the different walls serve for localization of sound sources, far-field recording and following speech processing. Besides video recording the personal web cameras mounted on the tables have internal microphones and are used for recording speech of each meeting participant.

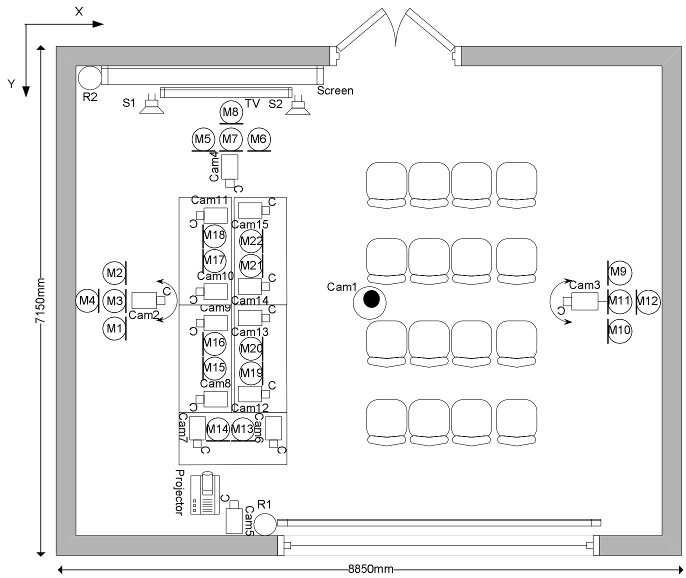


Fig. 1. Scheme of intelligent meeting room and equipment arrangement

A wide touchscreen plasma panel and multimedia projector (projection screen) are located one under another in the left side of the room and provide output of multimodal information. Operated electro gears are connected to the projection screen and curtain rail. The curtains are made from special lighttight cloth in order to suppress the outside influence on the illumination changing in the room. The processing of recorded audio-visual data, control the multimedia and electro mechanic equipment are performed by six four-cored computers, two multichannel audio boards Presound FirePod, as well as some devices providing cable and wireless network. The referred service equipment is mounted in a rack and located on the adjacent room from the meeting one. Thus, users inside the meeting room could see only appliances for input/output information, but other devices and computational resources are invisible. To provide service and the same time be hidden for a user is one of the main features of ambient intelligence.

Availability of multimodal user interface is a distinguishing characteristic of the developed intelligent meeting room. The developed earlier technologies of audio-visual data processing were successfully implemented in the room to provide natural and intuitively understandable way of interaction with the room equipment. Most important technologies are automatic speech recognition, speaker identification, sound source localization, detection of position and tracking moving objects and user faces, detection of user pose. A bimodal audio-visual Russian speech synthesis technology (talking head) is used for developed interactive applications.

A method for spectral-spatial analysis of speech activity has been developed. Spatial speaker localization is based on the calculation of phase difference between the signals recorded by different microphones. The energy level of mutual signal spectrum and estimation of acceptable position of a speaker is used for finding the boundaries of speech in a multichannel acoustical stream, recorded in noisy environments.

Russian speech recognition is realized by SIRIUS engine with a model of compact representation of extra large vocabularies based on two-level morphophonemic prefix graph (TMPG). Integration of morpheme and phonetic levels into a united tree-based structure of vocabulary provides compact representation of word-forms and their phonemic transcriptions [8]. Usage of the proposed graph while decoding continuous Russian speech provides formation of grammatically correct words at the recognition output and increase of recognition speed.

A model of text-independent speaker identification based on assessment of cepstral features distribution of the input speech signal was developed. The minimal length of the speech signal, necessary for speaker pattern training and decision making during verification, is thirty seconds. A multithreaded program model for speaker identification provides simultaneous estimation of several participants of a meeting.

Determination of position and face tracking are based on an algorithm for tracking movements of natural markers of human face. Automatic restoring of lost tracking points allows the face tracking system to increase robustness of identifying head-position during quick motions and occasional video noise. An algorithm for cursor controlling is adaptive to velocity of head movements. It provides comfortable cursor controlling on high-resolution screen, using a low-resolution camera. These algorithms have been embedded into the multimodal system ICANDO for assistance to disabled people at hands-free PC-controlling [6].

An algorithm for position and video tracking of moving object is based on estimation of difference between adjacent frames and takes into account the measurements of users, velocity and trajectory of their movements, acceptable regions of user appearance. It makes the algorithm robust to sudden changing luminosity and capable to differentiate moving objects closely located to each other. Intel OpenCV library is used for base procedures of video capture and processing. Also two statistic poses (staying and sitting) are detected in intellectual room applications. A criterion for pose recognition is height of an object. The list of recognized poses could be extended by some gestures, which users often apply during meetings [9]: nod, nutation, rise of a hand by sitting user, hand pointing.

A model of audio-visual speech synthesis (“talking head”) of Russian texts has been developed jointly with United Institute of Informatics Problems (Belarus) and University of West Bohemia (Czech Republic). Two methods for creation of visual synthesis of facial organs have been elaborated [10]: (1) model-based method; (2) data-based method. The former is used for creation of 3D avatars. The latter approach allows creating a 2D personalized audio-visual TTS-synthesis model. In this model visual facial synthesis is combined with acoustical speech synthesizer.

Figure 2 shows a technological framework of the intelligent meeting room. Collaborative work of the referred technologies supplies a room control system with data about current situation, participant's behavior and provides robust recognition of voice command due to the analysis of spatial-temporal, situational information and user preferences.

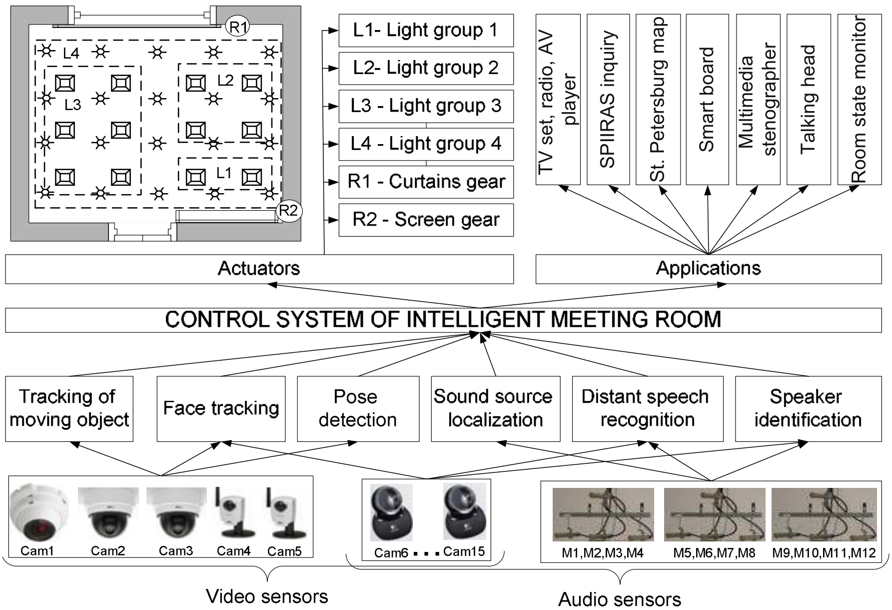


Fig. 2. Technological framework of the intelligent meeting room

Distant speech recognition based on microphone array processing allows a user to control light, curtains, projector screen, PTZ functions of video cameras and more complex devices like TV set, radio, audio-video player and other multimodal applications. In all the applications interactive feedback is realized by a talking head, which shows an awareness of the room about participant's behavior and pronounces required speech information.

3 Case Studies of Intelligent Meeting Room

First scenario, which was experimentally tested, was focused on conjugation of the actuators (light, curtains, and screen) with the distant speech recognition system and other manual controls. A user could give a command by voice, touchscreen panel, remote control panel or via web-interface from a computer or a mobile phone. A scheme of user command processing and possible automatic room control in accordance with a current situation is presented in Figure 3.

Data about current states of all the actuators, presence of users and their positions are recorded every time at the arrival of a user command. Boolean descriptions of

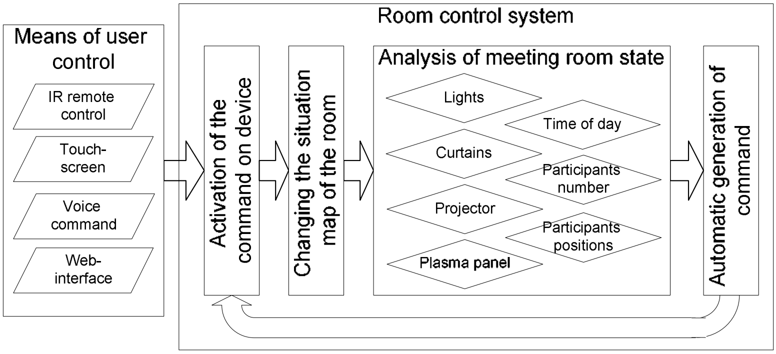


Fig. 3. Scheme of user and automatic control of the equipment of the room

actuator states were used for the light, the curtains, the screen, the plasma panel, the projector. The plasma panel and the projector have certainly many other functions, but their execution was realized without actuators in the framework of concrete applications. For example, changing graphical content showed by TV set and projector is processed by a computer, and then corresponding video signal outputs via DVI or VGA cables.

A description of situational information was simplified in given scenario too. Time of day was described by four categories (morning, day, evening, night). Number of participants presented inside the room was divided into four groups too (0, 1, 2-10, more than 10). Positions of participants were evaluated in relation to the meeting table, the plasma panel, the window. The chairs located in the right side of the room are used in the case of large number of presented participants. Therefore a special category of participant's position was defined: participants are distributed throughout the room. A situational map of the meeting room is used for verification of events and analysis of the room state. Table 1 presents structure and states of the map.

The black cells corresponding to the turned-on devices and other features of the room indicated a current situation. The right column consists of examples of the commands, automatically generated at occurrence of corresponding combination of events and equipment state. These commands were received during matching the real user command and the state of situational map. For some commands the conditions arisen them are slightly changed and adapted for implementation in automatic mode. For instance, a voice command to turn-off the light is said by user inside the room before she/he goes out. In automatic mode this action is performed when all the participants left the room. By this reason commands 6-8 in Table 1 to turn-off the light, lift the screen and close the curtains are performed when the number of participants is zero.

By Figure 3 now we can explain how to control the meeting room in the automatic mode. Changing the situation caused by user behavior or day time or switching the actuators calls the procedure of the map analysis. If a predefined situation is detected a corresponding command for activation or deactivation of actuators is sent. The interaction between various software modules distributed on several computers is accomplished by client-server architecture via TCP/IP protocol. At that command and notification of the actuator states are collected and processed in queue.

Table 1. The situational map of the meeting room and automatic commands

State of actuator devices								Situational information												Commands generated automatically, when specific combination of actuator states and events occurs
Light groups on								Time of day				Number of participants				Position of participants				
Group L1	Group L2	Group L3	Group L4	Curtains are closed	Screen is lowered down	Projector on	Plasma panel on	Morning	Day	Evening	Night	0	1	2-10	>10	All the participants are nearby the meeting table	There is a participant near the plasma/screen	There is a participant near the window	Participants are distributed throughout the room	
								■					■							1. Turn-on the light group L4
								■					■					■		2. Open the curtains
						■											■			3. Lower the screen
	■					■														4. Turn-off the light group L1
									■				■							5. Open the curtains
												■								6. Turn-off whole the light
												■								7. Lift the screen
												■								8. Close the curtains
																				…

Scenarios of control of multimedia appliances were based on special dialogues of speech interaction between a user and the meeting room. The multimodal applications "SPIIRAS inquiry" and "St. Petersburg map" were adopted based on similar systems realized in a multimodal kiosk [11]. Voice control system for TV set and radio implements commands to select a channel by its number or title, to change the settings of sound and picture. In the application "Smart board" voice commands intended for selecting color, width of pen, brush or other instruments for handwrited sketches on the touchscreen. Only examples of voice commands are mentioned above, which could be useful at interaction with the intellectual applications. Moreover, most of the commands could be activated by gestures on the touchscreen. In particular, in map-based applications a direct pointing to a graphical object is more needed, but speech commands are used for operation with objects [6].

Besides the current state of the dialogue the spatial position of a user should be taken into account. In contrast to control of the actuator devices the voice commands aimed for to the multimedia applications are perceived only from a space near the plasma panel (not far than 1.5 meters). This limitation helps to decrease number of false voice commands appeared as a result of background noise and parallel user conversations that increases accuracy of distant speech recognition.

Special attention should be paid to influence of activity of controlled devices during transitional stages on the systems of audio-visual data processing. For instance, the speech recognition and sound source localization systems should be notified before starting the gears of screen and curtains, because their noise influences on the system performance. The analogical situation is appeared at turning-on the multimedia projector and condition system. Since these devices will work continuously then a system of spatial filtration of sound signal should suppress noises before the speech recognition phase.

Opening the curtains can significantly influence on changing the luminosity in the room and lead to fail of the video tracking systems, for example loss of existent objects or appearance a false object in the room. The same problem happens when turning-on/turning-off the light, so the systems of video processing should be preliminary notified about changing the luminosity in the meeting room.

Verification of the technological framework and experimental detection of potential conflicts at the room control appeared owing to uncoordinated work of equipment or unpredicted behavior of users are conducted. Analysis of the extended situational map, as well as discursive information of real dialogues, personified user data will allow us to extract templates of behaviors and preferences of main groups of users, scenarios of man-machine interactions and most important commands, which should be automatically performed to facilitate efficiency of meetings and lectures in the intelligent room.

4 Conclusion

The developed intelligent meeting room is a distributed system with the network of intelligent agents (software modules), actuator devices, multimedia equipment and audio-visual sensors. The main aim of the room is providing of meeting or lecture participants with required services based on analysis of the current situation. Awareness of the room about spatial position of the participants, their activities, role in the current event, their preferences helps to predict the intentions and needs of participants. Context modeling, context reasoning, knowledge sharing are stayed the most important challenges of the ambient intelligent design.

Assignment of easy-to-use and well-timed services, at that stay invisible for user, is one of another important feature of ambient intelligent. In the developed intelligent room all the computational resources are located in the adjacent premises, so the participants could observe only microphones, video cameras, as well as equipment for output of visual and audio information. Implementation of multimodal user interface capable to perceive speech, movements, poses and gestures of participants in order to determinate their needs provides the natural and intuitively understandable way of interaction with the intelligent room.

Development of a network of intelligent meeting rooms gives the opportunity to organize a videoconference between spatially distributed participants and facilitates to increase collaboration, access to higher knowledge/competence, reduce costs for transport and staff, and increase the quality of education due to automatic immediate monitoring by every student during the lessons. Using the various combinations of multimodal interfaces and the equipment of the intelligent meeting room the fundamental issues of human-machine interaction are studied and applied models in security medicine, robotics, logistics and other scientific areas are investigated now.

Acknowledgments. The intelligent meeting room of SPIIRAS was supplied in 2008 at the financial support of the Russian Foundation for Basic Research (grants № 08-08-05004-б and № 08-08-01280-э_б). The models of speech and multimodal interfaces were developed in frameworks of the projects the Russian Foundation for Basic Research (grants № 07-07-00073-a, № 08-08-00128-a, № 08-07-90002-Бел_а) and SIMILAR NoE FP6-IST № 507609.

References

1. Ducatel, K., Bogdanowicz, M., Scapolo, F., Leijten, J., Burgelman, J.-C.: ISTAG - Scenarios of Ambient Intelligence in 2010. European Commission Community Research (2001)
2. Aldrich, F.: Smart Homes: Past, Present and Future. In: Harper, R. (ed.) Inside the Smart Home, pp. 17–39. Springer, London (2003)
3. Gann, D., Venables, T., Barlow, J.: Digital Futures: Making Homes Smarter. Chartered Institute of Housing, Coventry (1999)
4. Masakowski, Y.: Cognition-Centric Systems Design: A Paradigm Shift in System Design. In: 7th International Conference on Computer and IT Applications in the Maritime Industries, pp. 603–607 (2008)
5. Degler, D., Battle, L.: Knowledge management in pursuit of performance the challenge of context. Performance Improvement 39(6), 25–31 (2007)
6. Tzovaras, D. (ed.): Multimodal User Interfaces: From Signals to Interaction. Springer, Heidelberg (2008)
7. Chai, J., Pan, S., Zhou, M.: MIND: A Context-based Multimodal Interpretation Framework. Kluwer Academic Publishers, Dordrecht (2005)
8. Ronzhin, A.: Topological peculiarities of morpho-phonemic approach to representation of the vocabulary of Russian speech recognition. Bulletin of computer and information technologies, Moscow (9), 12–19 (2008)
9. Wallhoff, F., Zobl, M., Rigoll, G.: Action segmentation and recognition in meeting room scenarios. In: International Conference on Image Processing (ICIP 2004), pp. 2223–2226 (2004)
10. Lobanov, B., Tsirulnik, L., Železný, M., Krňoul, Z., Ronzhin, A., Karpov, A.: Audio-Visual Russian Speech Synthesis System. Informatics, Minsk, Belarus 20(4), 67–78 (2008)
11. Karpov, A., Ronzhin, A.: An Information Enquiry Kiosk with a Multimodal User Interface. In: 9th International Conference Pattern Recognition and Image Analysis (PRIA-2008), Nizhny Novgorod, Russia, pp. 265–268 (2008)

Context-Aware Team Task Allocation to Support Mobile Police Surveillance

Jan Willem Streefkerk, Myra van Esch-Bussemakers, and Mark Neerincx

TNO Human Factors, Kampweg 5,
3769 DE Soesterberg, the Netherlands
{j.w.streefkerk,myra.vanesch,mark.neerincx}@tno.nl

Abstract. To optimally distribute tasks within police teams during mobile surveillance, a context-aware task allocation system is designed and evaluated with end-users. This system selects and notifies appropriate team members of current incidents, based on context information (officer availability, officer proximity to the incident and incident priority) and decision rules. Eight teams of three experienced police officers evaluated this system in a surveillance task through a virtual environment, comparing it to a non-adaptive system. Task performance, communication, workload and preferences were measured. Results show that team communication, decision making and response times improve using the adaptive system and that this system is preferred. We conclude that context-aware task allocation helps police teams to coordinate incidents efficiently.

Keywords: Context-aware computing, mobile computing, police surveillance, task allocation, notification.

1 Introduction

To work efficiently as a distributed team, mobile police officers need to coordinate actions together. During surveillance, current incidents require fast and accurate responses from available team members. However, keeping track of availability and appropriately allocating tasks to team members is challenging in such a distributed work environment [1, 2]. In addition, unwanted interruptions can cause distraction [3], e.g. a colleague requesting assistance while you are talking to a violent suspect. This results in increased response times and potentially dangerous situations. What is needed is a system that supports team decision making and task allocation and provides appropriate notification to team members (e.g. on current incidents).

In this study, a support system is designed that provides advice on task allocation (which team member can handle which incident best) based on officer availability, task priority and officer proximity to the incident location. Selected team members are notified using appropriate notification styles (timing and presentation of notifications) to limit interruptiveness of notifications [4]. This team task allocation support is evaluated in a surveillance task with police teams, addressing the following questions: 1) how can user availability, task priority and proximity be used for team task allocation support and 2) what are the effects of this support on police team task performance, workload and subjective judgments? We expect that this support will improve

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 88–97, 2009.

task allocation and task performance, result in less team communication and positive user preferences. Based on this study, implications for the design of mobile professional support systems are discussed.

1.1 Previous Research on Task Allocation

Employing mobile technology in operational domains aims to increase shared situation awareness and enable flexible decision making, for example for soldiers (e.g. Network Centric Warfare; [5]) and first responders [6]. These efforts explicitly visualize team information in geographical overviews [2] or using mobile awareness cues [7]. Such designs for activity awareness in mobile computer-supported cooperative work (e.g. [8]) lead to increased team performance and awareness as well as reductions in mental workload [2, 9]. However, users still have to integrate information from a small mobile display, straining their cognitive resources. In addition, these awareness displays do not directly support task allocation.

To realize efficient task allocation support systems, such systems have to be aware of team members' activities and locations. Currently, mobile context-aware services adapt information presentation to dynamically changing user needs or changes in the work environment (e.g. [10]). Mobile context-aware information delivery was proposed for fire-fighters [11, 12] and construction workers [13]. In the police domain, an in-car support system was proposed to improve task allocation between the emergency room and police officers in the field, based on police officers' current tasks [14]. However, as no user evaluations were reported, it is not clear whether these systems actually support decision making and task allocation.

Using context information to optimize task allocation falls within the area of Augmented Cognition, which seeks to model the user and use context to dynamically adapt the user-system interaction. For example, in the naval domain, tasks were allocated dynamically to human operators or to an automated system based on task information (e.g. priority and number of radar contacts) or physiological measures assessing workload [15]. On the downside, some authors argue that such high levels of system automation are not advisable in time-critical environments [16]. Still, few empirical users studies address automated task allocation in mobile, operational domains [15, 17].

In the current study, a prototype system is designed that not only visualizes team information to support activity awareness, but also advises team members about appropriate task allocation. Although some studies have focused on task allocation and notification in the police domain [14, 18], such support has not yet been realized for mobile police officers. To assess how this task allocation support prototype effects police team performance, the prototype is evaluated with police end-users in a virtual environment. This allows users to experience the adaptive system within the task flow and facilitates control over external variables [17, 19].

2 Designing Team Task Allocation Support

Following a concise analysis of police team surveillance based on police interviews [2], a task allocation support system was designed and implemented. Based on a set of decision rules (on officer availability, incident priority and officer proximity), this system distributes incidents optimally over team members.

2.1 Police Team Surveillance

Police officers on surveillance work together as distributed, ad hoc teams. When an incident occurs, they communicate with colleagues to determine who should handle which incident. For this process to work efficiently, they have to be aware of location and priority of current incidents as well as location, identity and availability of colleagues.

Efficient task allocation is threatened by two problems. First, police officers currently have no overview of availability and location of team members. This makes it often unclear who is available to handle an incident, potentially resulting in miscommunications or incidents that remain unattended. Second, team members communicate using radio transceivers over an open channel. For officers who are handling a critical incident, not all communication is directly relevant and might cause unwanted interruptions. Busy police officers have been observed to turn off their radios. However, they still need to be aware of other high priority emergency situations (e.g. when colleagues are requesting assistance).

To address these problems, a mobile support system is designed that is aware of team members' location and availability (handling an incident or not) and the priority of the incident (high or low). This information is acquired from location tracking data, user responses and established incident categorizations in the police domain. Based on this knowledge and a set of decision rules, the system selects the most appropriate team member(s) to handle the current incident. These team members receive a notification message with task allocation advice (i.e. "John and Mary can handle the burglary incident best"). As in previous work [4], the presentation style of these messages (information density and auditory salience) is adapted to limit unwanted interruptions. The prototype uses the following decision rules:

1. *Priority*: high priority incidents require the nearest two available officers as soon as possible while low priority incidents require the nearest available officer.
2. *Availability*: if the nearest officers are busy with a lower priority incident, they should switch to the new incident. If they are busy with a higher priority incident, they should finish that incident first.
3. *Notification*: if officers are selected to handle an incident, the full incident message is presented with a salient notification sound. If they need to be aware that an incident is waiting for them, the system presents an indicator with a less salient sound. If they are not selected to handle the incident, an indicator is presented without sound.

2.2 Prototype Implementation

A prototype support system is implemented for experimental purposes using a simulated Personal Digital Assistant (PDA) on a touch screen monitor. It provides a geographical north-up map with icons indicating team members' location, identity (name) and availability (red icon means busy, green icon means available) as well as the

Fig. 1. PDA screenshots showing the geographical map with the officer's location (*left*), an incident message (*center*) and the task list (*right*)

location of incidents. The map is centered on the users' location and can be dragged to reveal the rest of the map. See Fig. 1 for screenshots of the application.

Incident messages are displayed as full text messages with two buttons to "Accept" or "Ignore" the incident. Accepted messages move to the task list and can be checked off when the incident is finished. User actions ("Accept", "Ignore", "Finish") are used to infer user availability. Indicators are presented as small clickable icons in the lower right corner of the screen, opening the incident message when clicked.

3 Evaluation

In this study, police teams performed a surveillance task through a virtual city environment. The task allocation support system presented low or high priority incident messages. At these moments, team members negotiated who would handle which incident, navigated to the incident location and handled the incident. Task allocation advice, notification presentation and communication were manipulated, creating two conditions (adaptive and control). Effects on task performance, workload and subjective ratings were assessed between the two conditions.

3.1 Method

Participants. Eight teams of three police officers (20 male, 4 female, mean age = 33.0 years, SD = 9.9) participated in this study. All team members were experienced police officers (average 11.2 years of experience) and had collaborated previously with each other on surveillance. They used personal computers on a daily basis and only two teams used a PDA for police work.

Surveillance Task and Incident Handling Task. Teams performed the surveillance task and incident handling task through a virtual city environment [20]. The surveillance task required them to collect a maximum of 30 targets, represented by barrels that appeared at random locations throughout the environment.

The incident handling task was a time-paced, scenario-based task. At predetermined moments during the surveillance round, the system presented in total twelve incident messages to the team. Six incident messages indicated high priority incidents, which had to be handled by two colleagues together. The other six indicated low priority incidents, which could be handled by a single team member. Team members suspended the surveillance task to read the incident message and communicated with colleagues (using a headset) who would handle this incident. The selected team member(s) responded to it (using the “Accept” or “Ignore” buttons below the message) and navigated to the incident location as fast as possible (see Fig. 2). Handling the incident consisted of reading and memorizing the incident description on screen. When done, they checked the incident off the task list and returned to the surveillance task. Participants could decide for themselves when to attend to each message, whether or not to accept an incident and which of their colleagues to approach for assistance.

Experimental Design. A within-subjects design was employed with two experimental conditions (adaptive or control). In the adaptive condition, the system provided task allocation advice and adaptive notification following the decision rules. Team members could choose to communicate with all team members or selected team members only (closed channel). In the control condition, full incident messages were presented to all team members without task allocation advice and the communication channel was open to all team members.

Two similar experimental scenarios (equal duration, number and type of incidents) were established in cooperation with two experienced police officers to maximize external validity. All teams experienced both conditions and the presentation order of the conditions and scenarios was counterbalanced across teams to avoid order effects.

Measures. Before the experiment, age, gender, (mobile) computer experience and police experience were assessed using a questionnaire. Furthermore, spatial ability was assessed in a computer-based spatial rotation task [21].

During the experimental sessions, task performance on the surveillance task was measured as the *total distance traveled* and the number of *targets* collected. Task performance on the incident handling task was measured as the *response time* to incident messages, *errors* in decision making, *incident handling time*, *total time on task* and recall of incident *details*. Furthermore, the number of *communication utterances* on task allocation between team members was counted. These variables were measured per team and averaged over incidents.

Subjective judgments were collected individually using questionnaires and rating scales. After each session, experienced *workload* was measured using the RSME [22] and judgment on *own performance* and *team performance* was measured with a six-item team effectiveness scale. After both sessions, team members were asked individually to compare both experimental sessions. On the *preference* questionnaire they indicated which of the two prototypes they would prefer in their daily police practice regarding task allocation advice, presentation of the messages and team communication.

Fig. 2. The virtual environment with an incident location (*left*) and a police officer behind the experimental setup (*right*)

Apparatus. Participants were seated behind two 17” monitors, one above another. The top monitor displayed the virtual environment and the incident details. Participants moved through the environment using a game controller. The bottom (touchscreen) monitor displayed the simulated PDA and communication interface (see Fig. 2). To avoid overhearing each other, city background noise was played over the headset. While navigating through the environment, the PDA was blanked out to avoid overreliance on the geographical map.

Procedure. In total, the experiment took about three hours to complete. First, the personal characteristics questionnaire and the spatial ability test were administered. Participants received instructions on both tasks and familiarized themselves with navigation and incident handling in two short practice scenarios (control first, adaptive second). In the control condition, participants were instructed to follow the set of decision rules for task allocation (see paragraph 2.1), while in the adaptive condition the system provided task allocation advice. The two experimental sessions took about twenty minutes each, after which the RSME, the performance and detail recall questionnaires were administered. After both sessions, the preference questionnaire was administered.

3.2 Results

Data on all performance variables was averaged and compared per condition using t-tests for repeated measures. Means for all variables are presented in Table 1. For response time, decision errors and navigation efficiency, follow-up analyses per priority level (high or low) were performed. Subjective judgments were analyzed using non-parametric tests. Multiple regression analyses were performed on performance measures, communication and workload with age, spatial ability, education, computer experience, police experience and game experience (averaged over teams) as predictor variables.

Surveillance Task Performance. The difference in total *distance* traveled between de adaptive and control condition approached significance ($t(7) = 2.13$, $p = 0.07$). Less distance was traveled in the adaptive condition. On average, more *targets* were collected in the adaptive condition (M = 18.5) compared to the control condition (M = 17.4). However, this difference was not significant ($t(7) = -0.44$, $p = 0.67$).

Regression analysis showed that variance in distance traveled was significantly predicted by age (R^2 adj. = 61%, B = 8956, $p < 0.05$) and variance in targets collected was also explained by age (R^2 adj. = 64%, B = -0.5, $p < 0.05$); younger teams collected more targets and traveled less distance in the control condition. In the adaptive condition, no significant predictors were found on these variables.

Incident Handling Task Performance. *Response time* to incident messages was slightly lower for the adaptive condition, however not significant. This can be explained by the extra line of message text (with the task allocation advice) that had to be read in this condition. When response times were analyzed separately for high and low priority messages, the interaction effect of condition and priority approaches significance ($F(1, 7) = 4.32$, $p = 0.076$; see Fig. 3). In the control condition response time to low and high priority incidents is almost identical, while in the adaptive condition, participants' response time differs between low and high priority incidents.

Both *incident handling time* and *total time on task* did not differ significantly between the adaptive and the control condition. Regression analysis showed that variance in incident handling time was also predicted by age (R^2 adj. = 91%, B = -5.37, $p < 0.01$) and variance in time on task in the control condition was predicted by age (R^2 adj. = 65%, B = -4.92, $p < 0.05$); younger teams took more time than older teams. This effect was not present in the adaptive condition.

The number of *decision errors* on task allocation was lower in the adaptive condition (M = 3.4) than in the control condition (M = 5.0), approaching significance ($t(7) = 2.09$, $p = 0.07$). Analyzed separately for high or low priority incidents, no significant interaction effect was found (see Fig. 4). The adaptive support helped teams to reduce decision errors.

The number of *details recalled* was slightly higher in the adaptive condition than in the control condition (see Table 1), however not significant. Regression analysis showed that variance in detail recall in the control condition was predicted by age (R^2 adj. = 59%, B = -0.34, $p < 0.05$); older teams recalled less details. However, in the adaptive condition, this effect was not present.

Communication. The number of *communication* utterances on task allocation differed significantly between conditions ($t(7) = 4.17$, $p < 0.005$). In the adaptive

Table 1. Means on the main performance variables (TG = targets, RT = response time, DE = decision errors, HT = incident handling time, TT = total time, Com = communication utterances, Det = details recalled and WL = workload) for the control (*Co*) and adaptive (*Ad*) conditions. * significant at p < 0.05, **bold** indicates a trend approaching significance.

	Distance	TG (#)	RT (s)	DE (#)	HT (s)	TT (s)	Det (#)	Com (#)	WL
Co	**383425**	17.4	11.0	**5.0**	70.1	129.5	16.9	***33.2**	49.3
Ad	**332734**	18.5	13.5	**3.4**	67.0	123.6	17.5	***23.3**	51.0

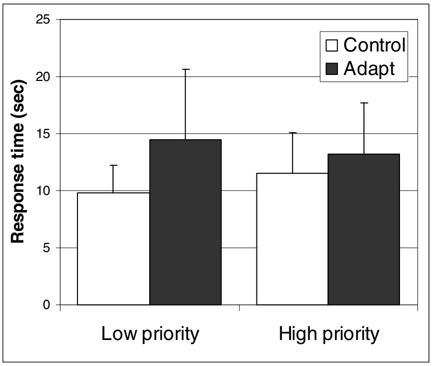


Fig. 3. Response times to low and high priority incident messages

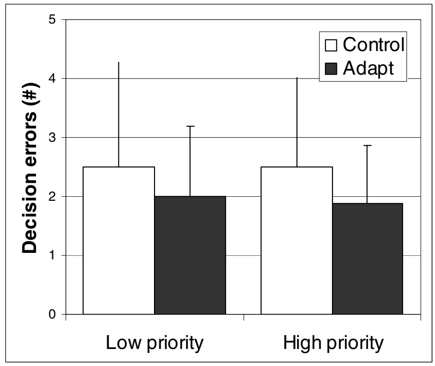


Fig. 4. Number of decision errors on low and high priority incidents

condition, team members communicated less on task allocation than in the control condition (23 and 33 utterances respectively). When they had the choice between open or closed channel of communication in the adaptive condition, informal observations showed that almost all teams preferred and used an open channel.

Workload. There was no significant difference in *workload* between both conditions. Regression analysis showed that workload in the control condition was predicted by spatial ability (R^2 adj. = 89%, B = -5.80, $p < 0.05$), game experience (B = 14.8, $p < 0.05$) and education (B = 8.06, $p < 0.05$); participants with high spatial ability and less game experience indicated lower workload ratings. Workload in the adaptive condition was also predicted by spatial ability (R^2 adj. = 84%, B = -5.17, $p < 0.05$) and game experience (B = -8.18, $p < 0.05$), but showed that participants with *more* game experience indicated lower workload.

Subjective Judgments. The questionnaire items on own performance and team performance showed no significant differences between conditions. Participants did not rate their own performance or team performance differently in one of the conditions. The mean scores on the team effectiveness scale showed a ceiling effect (5.8 and 5.9 for control and adaptive condition respectively).

Participants' *preferences* after both conditions showed that 76% of the participants preferred the adaptive condition in their daily police work because it supported decision making. Half of the participants preferred the adaptive condition because of the lower disruptiveness of messages. However, 58% of the participants found it to be more difficult to divide attention between the PDA and the surveillance in the adaptive condition.

4 Discussion and Conclusion

This study evaluated team task allocation support based on relevant context factors (location, availability and priority). In a surveillance task with experienced police teams, two conditions (with and without context-aware task allocation and notification) were compared. Using task allocation support, less team communication and less decision errors are observed and less distance is traveled. In addition, adaptive notification causes response times to be more varied, appropriate for the priority of the incident. The majority of the officers preferred this support in their daily work, although some found the adaptive system behavior hard to understand. Regression analysis showed that older police officers profited from the support in terms of more details recalled, distance and targets collected and younger officers profited in terms of incident handling time. Our results show that context-aware task allocation support helps police teams in decision making and communication.

Contrary to our expectations, no effects of support were found on time on task, incident handling time and workload; the support does not make police officers faster nor lessen their workload. The time benefits of the task allocation support may be too small compared to the incident handling time (over 120 seconds). In addition, learning to work with an adaptive system might have increased officers' workload. These effects are expected to decrease with prolonged system use. An interesting observation on team collaboration is that without task allocation support, tasks were allocated to whoever called first or loudest. While this may not have been the most appropriate decision, still teams rated team performance very positively.

These results have implications for the design of task allocation support systems on mobile devices. Because the use context (location, availability) can change unexpectedly, task allocation advice may become outdated or wrong. Consequently, the task list must allow team members to pass over incidents or tasks to others. Or task allocation should be dynamically revised, based on the current situation. This is an opportunity to extend the principles of Augmented Cognition to the mobile domain. Further research should focus on the usability and predictability of such dynamic task allocation systems. In ongoing research, we will investigate how teams deal with unexpected breakdowns in task allocation support and what the role is of shared situation awareness in handling such situations.

Acknowledgements. The MultimediaN project is funded by the Dutch Ministry of Economic Affairs. We thank Bert Bierman, Peter Duin of VTS Politie Nederland, Roland Smeenk, Nanja Smets, Tjerk de Greef and all police participants for their contributions to this study.

References

1. Ferscha, A.: Workspace awareness in mobile virtual teams. In: Proc. of WET ICE 2000, Gaithersburg, MD, USA, pp. 272–277 (2000)
2. Streefkerk, J.W., Wiering, C., Esch-Bussemakers, M., Neerincx, M.: Effects of Presentation Modality on Team Awareness and Choice Accuracy in a Simulated Police Team Task. In: Proceedings HFES, vol. 52, pp. 378–382. HFES, New York City (2008)
3. Jameson, A.: Modeling Both the Context and the User. Pers. & U. Comp. 5, 29–33 (2001)
4. Streefkerk, J.W., Esch-Bussemakers, M., Neerincx, M.: Context-Aware Notification for Mobile Police Officers. In: Harris, D. (ed.) HCII 2007 and EPCE 2007. LNCS, vol. 4562, pp. 436–445. Springer, Heidelberg (2007)

5. Adkins, M., Kruse, J., Younger, R.: Ubiquitous computing: omnipresent technology in support of Network Centric Warfare. In: Proc. of HISS, pp. 551–559 (2002)
6. Smets, N., Te Brake, G., Lindenberg, J., Neerincx, M.: Influence of mobile map size and user capacities on situation awareness tested in a virtual environment. In: Proc of ISCRAM 2007, pp. 557–564 (2007)
7. Oulasvirta, A., Petit, R., Raento, M., Tiitta, S.: Interpreting and Acting on Mobile Awareness Cues. Human-Computer Interaction 22, 97–135 (2007)
8. Carroll, J.M., Rosson, M.B., Convertino, G., Ganoe, C.H.: Awareness and teamwork in computer-supported collaborations. Int. with comp. 18, 21–46 (2006)
9. Kaber, D., Endsley, M.: The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. Th. Iss. in Erg. Sc. 5, 113–153 (2004)
10. Dey, A., Abowd, G., Salber, D.: A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications. Hum.-Comp. Int. 16, 97–166 (2001)
11. Jiang, X., Chen, N., Hong, J., Wang, K., Takayama, L., Landay, J.: Siren: Context-aware Computing for Firefighting. In: Proc. of Pervasive, Vienna, Austria (2004)
12. Camp, P., Hudson, J., Keldorph, R., Lewis, S., Mynatt, E.: Supporting Communication and Collaboration Practices in Safety-Critical Situations. In: Proc. CHI, pp. 249–250 (2000)
13. Aziz, Z., Anumba, C., Ruikar, D., Carrillo, P., Bouchlaghem, N.: Intelligent wireless web services for construction—A review of the enabling technologies. Automation in Construction 15, 113–123 (2006)
14. Nulden, U.: Investigating police patrol practice for design of IT. In: Proc. of CHI 2003, pp. 820–821 (2003)
15. Arciszewski, H., de Greef, T., van Delft, J.: Adaptive Automation in a Naval Combat Management System. IEEE SMCA (in press)
16. Cummings, M.: Automation Bias in Intelligent Time Critical Decision Support Systems. In: Proc. of AIAA Intelligent Systems, Chicago, IL, USA (2004)
17. Weibelzahl, S.: Problems and Pitfalls in Evaluating Adaptive Systems. In: Chen, Magoulas (eds.) Adaptable and Adaptive Hypermedia Systems, pp. 285–299. IRM (2005)
18. Streefkerk, J.W., Esch-Bussemakers, M., Neerincx, M.: Field evaluation of a mobile location-based notification system for police officers. In: Proc. of MobileHCI, pp. 101–108. ACM, Amsterdam (2008)
19. Streefkerk, J.W., Esch-Bussemakers, M., Neerincx, M., Looije, R.: Evaluating Context-Aware Mobile Interfaces for Professionals. In: Lumsden, J. (ed.) Handbook of Research on User Interface Design and Evaluation for Mobile Technology, pp. 759–779 (2008)
20. Te Brake, G., de Greef, T., Lindenberg, J., Rypkema, J., Smets, N.: Developing adaptive user interfaces using a game-based simulation environment. In: Proc. of ISCRAM (2006)
21. Neerincx, M., Lindenberg, J.: Supporting individual situation awareness in web-environments. In: Pikaar, R. (ed.) Ergonomie in Uitvoering, pp. 144–155. NVvE (1999)
22. Zijlstra, F.: Efficiency in work behavior. A design approach for modern tools. Delft University of Technology (1993)

Operational Brain Dynamics: Data Fusion Technology for Neurophysiological, Behavioral, and Scenario Context Information in Operational Environments

Don M. Tucker and Phan Luu

Electrical Geodesics, Inc., Eugene, Oregon
NeuroInformatics Center, University of Oregon, Eugene, Oregon
dtucker@egi.com

Abstract. Classical laboratory studies of human performance have always required some form of data integration, such as the synchronization of stimulus display, behavioral accuracy, and reaction time. Studies of performance in operational environments have typically been limited in the precision of behavioral observations. As improved digital informatics have expanded the laboratory data acquisition from a few bytes to terabytes, there has been a similar expansion in both the opportunities and the challenges for data fusion.

Keywords: EEG, information systems, brain activity, neuroergonomics.

1 Introduction

An advanced window on human neurophysiological function has been opened by dense-array (256-channel) electroencephalography (DA-EEG). The improved sampling of the brain's electrical fields has been combined with improved physics models of the human head to allow accurate estimates of the electrical source activity of specific brain networks, such as the ventromedial frontal cortex or posterior cinguate cortex, that are known to be required for effective attention and cognition in demanding military environments. Recent advances in dermal bond hydrogel technology have improved DA-EEG signal quality even in high noise high movement operational environments such as mounted vehicle platforms. Advances in computer vision have allowed inobtrusive capture of critical details of behavior, such as head and eye tracking, in operational as well as laboratory environments with the millisecond accuracy required for fusion with electrophysiological data. At least in simulator environments, the instrumentation of the simulator software allows precise timing and description of events in the simulated operational context. With improved video recognition and sensor technologies, the information on operational contexts is also expanding. Advances in high-performance computation have allowed powerful mathematical algorithms such as independent components analysis and directed components analysis to separate unique sources of variance in the fused data streams. We describe a net-centric, distributed-parallel informatics architecture for increasing the bandwidth of the instrumentation and fused analysis of neurophysiological, behavioral, operational scenario events.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 98–104, 2009.

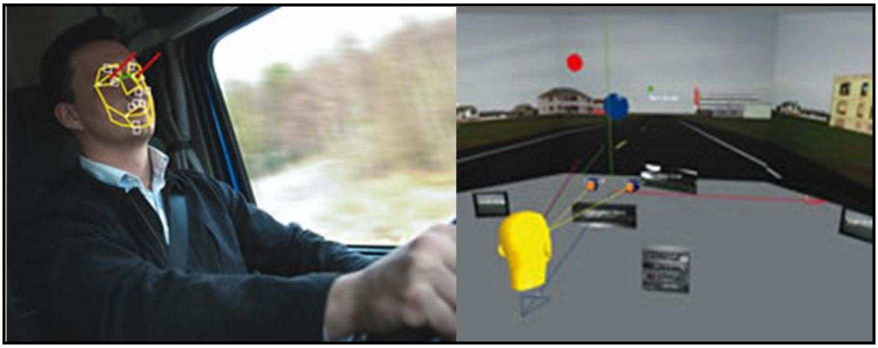


Fig. 1. Tracking of a subject's gaze with free head movement in a 3D environment

The immediate objectives of the proposed project are to: 1) implement a mobile, field-deployable hardware and software platform (AmpServer) capable of integrating and synchronizing the acquisition of neurological (high-density electroencephalography and near-infrared spectroscopy), behavioral (head- and eye-tracking), and autonomic (e.g., EKG) data during field operations, and 2) adapt and refine advanced artifact-cleaning and pattern classification methods to identify and separate the relevant data signals.

Under the DARPA Augmented Cognition program, EGI developed methods for real-time data acquisition and analysis to integrate dense array (256-channel) EEG with head and eye tracking information from infrared video. Under the DARPA/NGA Neurotechnology for Intelligence Analysts program, EGI developed methods for real-time recognition of visual system responses that indicate that an analyst has detected a military target in a rapid (10/sec) stream of visually presented satellite images. Under the ONR Human Performance Training and Education program, EGI has developed methods for assessing the neural mechanisms in the development of expertise during training.

Learning, or performance, seen as action regulation inherently emphasizes the need to adjust behavior according to both internal states and external demands, requiring different learning and memory systems; these systems reflect cybernetic constraints on action control. Learning, adaptive performance, and memory naturally arise from these action regulation processes. Two complementary cortico-limbic-thalamic circuits have been identified, each providing a unique strategic control on the learning process [1]. The ventral limbic circuit is made up of the anterior cingulate cortex (ACC) and the medial nuclei of the thalamus, with input from the amygdala. This ACC-based circuit is triggered by exogenous feedback, and leads to rapid changes in learning in response to new information, discrepancies with expectations, and threat. It is involved in the early stages of learning, whenever new tasks must be learned, or when routine actions and a priori knowledge are no longer appropriate for current demands [2-4]. The dorsal limbic circuit is centered on the posterior cingulate cortex and anterior ventral nucleus of the thalamus, with input from the hippocampus. It is involved in the later stages of learning and expert performance [5], when consolidation of information into long-term memory is important [2]. In these late stages, a contextual model is fully formed, and minor changes that are consistent with the contextual model can be made with minimal attention demands.

Human research in our laboratory with DA-EEG measures has yielded results largely consistent with this model. For example, we have observed greater ACC activity, as assessed through source analysis of the scalp-recorded EEG, when expectancies were violated [6], under particularly challenging performance demands [7,8], and following errors and negative feedback [9]. Moreover, increased anxiety was associated with the modulation of ACC engagement [10]. In contrast, we observed greater activity associated with the PCC circuit during the later stages of learning [11], after extensive practice [7,8], and in expert versus novice performance. Extension of these findings to realistic operational environments will help in identifying individual differences and contextual events that impact these fundamental self-regulatory mechanisms and enhance or impede adaptive performance.

2 Method

Significant technological advances have been made in the field of eye tracking. This project uses the state-of-the-art Smart Eye eye-tracking system. The Smart Eye system is completely unobtrusive (i.e., remote and not head mounted), and measures eye and head movements given the inputs from up to four cameras at 60 fps (see Figure 1). Head movement is measured to an accuracy of 0.5 degrees (rotation). The accuracy of the computed gaze-vector is 1 degree. The Smart Eye system allows us to track saccades and fixation within a 210-degree field of view. Pupillometry and a video stream with the image analyst's gaze position overlaid on the scene camera video, are also available as output. Additional benefits include flexible camera-mount positions, fast camera calibration, and handling of occluded cameras.

EGI, in collaboration with Smart-Eye engineers, has integrated the Smart-Eye system into our EEG acquisition platform. Because of the unobtrusive nature of this technology as well as the importance of understanding behaviors during performance we propose to employ it for tracking of attentional focus in a complex field environment. Integration of data acquisition from multiple sensors can be enabled through use of a common, network-capable, software architecture. Currently, EGI employs a software application called AmpServer for our dense-array EEG system. AmpServer has the capability to control multiple amplifiers, if they are all connected to the same machine and the bandwidth for all amplifiers are within the limits of Firewire technology. If the bandwidth requirements exceed the limits and the application requires integration of multiple amplifiers on the same machine, then multiple Firewire cards can be utilized. AmpServer currently is developed on Mac OS X but can be modified to run on Linux or Vista (when it is stable) with minor to moderate work. AmpServer can be made to support non-EGI amplifiers provided the amplifiers are stable and control and interface protocols are documented (see Figure 2).

With AmpServer as the platform, anyone can write client applications (on any platform) to access the raw data being broadcast by AmpServer. Alternatively, NetStation (EGI's acquisition software) can be used as the client.

The importance of electroencephalography (EEG) for tracking a human operator's cognitive state is well established. Moreover, our understanding of real-time brain

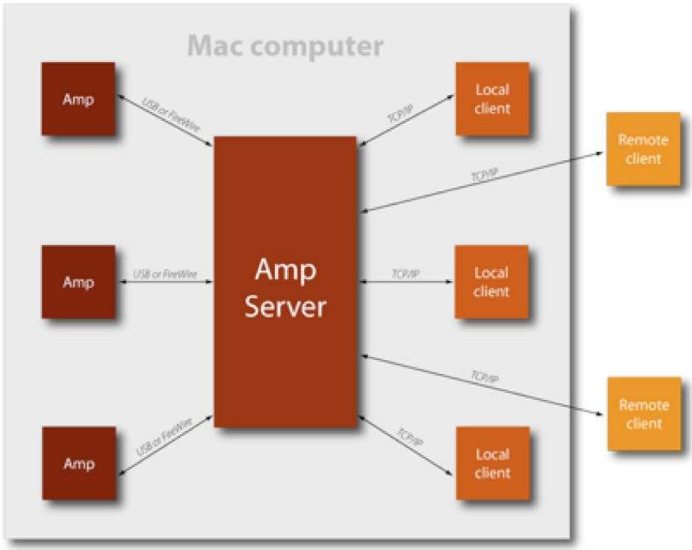


Fig. 2. Schematic of AmpServer architecture

activity is crucial in fulfilling the goal of monitoring performance to facilitate the mitigation of cognitive bottlenecks through dynamically modifying system behavior. Non-brain activity in continuous EEG severely masks and hampers the detection and interpretation of brain activity. Sources of non-brain activity include physiological artifacts, electromagnetic interference, and amplifier noise. The effectiveness of metrics derived from EEG to measure cognitive performance is severely diminished given the multitude of these artifact and noise sources. It therefore becomes both a crucial necessity as well as a major challenge to parse brain activity from the raw EEG signal in real time, while carefully minimizing the distortion of the actual brain activity components.

In a related effort (Luu et al., this volume), we developed a framework for detecting and extracting physiological artifacts due to ocular (i.e. eye blinks and movements) and cardiac activity from the recorded EEG in real time. Within this framework, the integration of continuous electrocardiography (EKG) as well as head and eye tracking enhances the robustness and stability of the artifact removal procedure. Both EKG and head and eye tracking have become ubiquitous in operator performance measurement environments.

Although Independent Component Analysis (ICA) has demonstrated the ability to cleanly separate ocular from brain activity, its reliance on computationally intensive higher-order statistics precludes its use in real-time applications. These higher order statistics can only be reliably calculated on long epochs, and exposes another drawback of ICA in that it is assumed that the measured EEG is derived from a limited set (equal to the number of EEG sensors) of spatially stationary brain and artifact generators over the entire epoch. Methods based on Principal Components Analysis (PCA), employing only computationally simpler second–order statistics, can be applied for artifact removal in real time. Special care is needed to ensure their effectiveness, as the reliance of PCA on orthogonal topographies has an important drawback. Existing PCA artifact removal methodologies can be classified as either methods that remove artifacts without considering brain activity, or techniques that attempt to separate artifact and brain activity. As part of our artifact removal framework, we propose a hybrid method that harnesses our ability to monitor and evaluate the temporal evolution of artifact activity. By identifying, selecting and segregating time slices of EEG data from contaminated and artifact-free epochs, we derive separate, finely detailed topographies for the artifact and brain activity in the signal, enabling a much cleaner removal of artifact contamination without distortion of the brain activity measurements. The integration and synchronization of head and eye tracking with EEG acquisition is essential for extracting eye (and head) movement artifacts effectively. By employing a separate EKG trace, we can cleanly extract cardiac artifacts, even in the presence of spike activity emanating from brain sources.

3 Results

In a series of eight studies investigating the effects of stress in simulated flight and a task analogous to pilots executing instructions from air traffic control we found that many people experienced stress-induced decrements in performance, particularly as task difficulty increased. However, others experienced no ill-effects of stress and for some people performance actually improved, despite equal levels of task difficulty. A unitary arousal model or the Yerkes-Dodson quantitative model [12], which associated performance with levels of difficulty and arousal, thus cannot explain these differences in performance under stress. Instead, we observed that *qualitative* differences in the emotional response to stress best accounted for these findings. Our results indicated that several factors influenced emotion response to stress. The following are particularly important to our understanding of the effects of stress in operational environments:

1. Context: Predictability of the stressor was related to decreased anxiety.
2. Experience: Early exposure to a stressor (i.e., when learning a new task) was related to increased anxiety, larger stressor-condition performance decrements, and lower levels of competency after two weeks of training.
3. Appraisal: Both experimental manipulation and participants' own appraisals of the stressor were predictably related to emotion response to stress.
4. Trait differences: When emotion responses to stress were most variable (due to differences in predictability of the stressor and appraisal manipulations), trait anxiety and behavioral inhibition were predictive of emotion responses to stress, but did not directly predict stressor-condition performance.

Although the body is adapted to respond with little or no ill effect to the acute mobilization of physiological distress reactions, it is clear that chronic or repeated activation of threat systems can have adverse long-term physiological, cognitive, and affective health effects [13]. Over the short-term, such reactions can also be maladaptive when individuals fail to flexibly regulate threat systems in the face of changing circumstances (e.g., when the threat no longer exists) or when the situation precludes fight or flight (e.g., work environments). Assessment of the autonomic stress response in complex, operational environments via EKG and EMG sensors in an integrated platform will provide a greater understanding of how these autonomic changes interact with the engagement of neural self-regulatory systems, such as the anterior, ventral limbic system under perceived threat. Head and eye-tracking measures will further clarify the attentional response to stress by tracking gaze duration and eye fixations. This information can indicate, for example, if one performer is more easily distracted under stress, rapidly shifting fixations across irrelevant information, whereas another performer is able to shift attention systematically to relevant information in a goal-directed fashion.

4 Discussion

The implementation of an integrated information environment for both behavioral and electrophysiological observations allows novel approaches to real-time measurement of human brain activity in operational environments. Key features of this implementation are single-trial data measures (rather than averaged event-related potentials) and exact precision of timing of high-bandwidth data streams. The technical capacities now available with video head and eye tracking and dense array EEG are well suited to the challenges of neuroergonomics in operational environments.

References

1. Gabriel, M., Burhans, L., Talk, A., Scalf, P.: Cingulate Cortex. In: Ramachandran, V.S. (ed.) Encyclopedia of the Human Brain, pp. 775–791. Elsevier Science, Amsterdam (2002)
2. Gabriel, M.: Functions of Anterior and Posterior Cingulate Cortex During Avoidance Learning in Rabbits. Prog. Brain Res. 85, 467–483 (1990)
3. Gabriel, M., Sparenborg, S.P., Stolar, N.: An Executive Function of the Hippocampus: Pathway Selection for Thalamic Neuronal Significance Code. In: Isaacson, R.L., Pribram, K.H. (eds.) The Hippocampus, pp. 1–39. Plenum, New York (1986)
4. Poremba, A., Gabriel, M.: Amygdalar Efferents Initiate Auditory Thalamic Discriminative Training-Induced Neuronal Activity. J. Neurosci. 21, 270–278 (2001)
5. Keng, E., Gabriel, M.: Hippocampal Modulation of Cingulo-Thalamic Neuronal Activity and Discriminative Avoidance Learning in Rabbits. Hippocampus 8, 491–510 (1998)
6. Frishkoff, G.A., Tucker, D.M., Davey, C., Scherg, M.: Frontal and Posterior Sources of Event-Related Potentials in Semantic Comprehension. Brain Res. Cogn. Brain Res. 20(3), 329–354 (2004)
7. Poulsen, C., Luu, P., Davey, C., Tucker, D.: Dynamics of Task Sets: Evidence from Dense-Array Event-Related Potentials. Cog. Brain Res. 24, 133–154 (2005)
8. Poulsen, C., Luu, P., Tucker, D., Scherg, M., Davey, C., Frishkoff, G.: Electrical Brain Activity during Task Switching: Neural Source Localization and Underlying Dynamics of Scalp-Recorded Potentials. Poster presented at the Tenth Annual Meeting of the Cognitive Neuroscience Society, New York, NY (March 2003)
9. Luu, P., Tucker, D.M., Derryberry, D., Reed, M., Poulsen, C.: Electrophysiological Responses to Errors and Feedback in the Process of Action Regulation. Psychol. Sci. 14(1), 47–53 (2003)

10. Poulsen, C., Luu, P., Crane, S.M., Quiring, J., Poolman, P., Tucker, D.M.: Neural Mechanisms of Self-Evaluation in Anxiety and Depression (in preparation)
11. Luu, P., Tucker, D.M., Stripling, R.: Neural Mechanisms for Learning Actions in Context. Brain Res. 1179, 89–105 (2007)
12. Yerkes, R.M., Dodson, J.D.: The Relation of Strength of Stimulus to Rapidity of Habit-Formation. J. Comparat. Neurol. Psychol. 18, 459–482 (1908)
13. McEwen, B.S.: The Neurobiology of Stress: From Serendipity to Clinical Relevance. Brain Res. 886, 172–189 (2000)

Characterizing Cognitive Adaptability via Robust Automated Knowledge Capture

Robert G. Abbott and J. Chris Forsythe

Sandia National Laboratories, P.O. Box 5800, Albuquerque, NM 87185, USA
{rgabbot,jcforsy}@sandia.gov

Abstract. Applications such as individually tailored training and behavior emulation call for cognitive models tailored to unique individuals on the basis of empirical data. While the study of individual differences has been a mainstay of psychology, a prevailing assumption in cognitive theory and related modeling has been that cognitive processes are largely invariant across individuals and across different conditions for an individual. Attention has focused on identifying a universally correct set of components and their interactions. At the same time, it is known that aptitudes for specific skills vary across individuals and different individuals will employ different strategies to perform the same task [3]. Moreover, individuals will perform tasks differently over time and under different conditions (e.g. Taylor et al, 2004). To reach their full potential, systems designed to augment cognitive performance must thus account for such between- and within-individual differences in cognitive processes. We propose that cognitive adaptability is a trait necessary to explain the inherently dynamic nature of cognitive processes as individuals adapt their available resources to ongoing circumstances. This does not imply a "blank slate;" humans are predisposed to process information in particular ways. Instead, we assert that given variation in the structure and functioning of the brain, there exists inherent flexibility that may be quantified and used to predict differences in cognitive performance between individuals and for a given individual over time. This paper presents an early report on research we are undertaking to discover the dynamics of cognitive adaptability, with emphasis on a task environment designed to evoke and quantify adaptation in controlled experiments.

1 Cognitive Adaptability in Cognitive Modeling

While the study of individual differences has been a mainstay of psychology, a prevailing assumption in cognitive theory and related modeling has been that cognitive processes are largely invariant across individuals and across different conditions for an individual. Attention has focused on identifying a universally correct set of components and their interactions. Between-subject and within-subject variability is generally regarded as measurement error.

At the same time, it is known that aptitudes for specific skills vary across individuals and different individuals will employ different strategies to perform the same task [e.g. 3]. Moreover, individuals will perform tasks differently over time and under different conditions (e.g. Taylor et al, 2004). To reach their full potential, systems

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 107–113, 2009.

designed to augment cognitive performance must thus account for such between- and within-individual differences in cognitive processes.

We propose that *cognitive adaptability* is a trait necessary to explain the inherently dynamic nature of cognitive processes as individuals adapt their available resources to ongoing circumstances. This does not imply a “blank slate;” humans are predisposed to process information in particular ways. Instead, we assert that given variation in the structure and functioning of the brain, there exists inherent flexibility that may be quantified and used to predict differences in cognitive performance between individuals and for a given individual over time.

2 Testing and Characterizing Cognitive Adaptability

A fundamental challenge in establishing cognitive adaptability is modeling individuals’ relative strengths and weaknesses, and tendencies to adopt different strategies. Unfortunately, tools that permit human knowledge and behavior to be automatically modeled at a level of individual specificity have largely been ignored within the cognitive neurosciences. Automated Knowledge Capture (AKC) is the most promising avenue for efficiently supplying cognitive models tailored to differences relevant to performance, decision making, and learning in complex environments.

Sandia National Laboratories, the University of Memphis, the University of Notre Dame, and the Mind Research Network are undertaking a study to test two foundational hypotheses of cognitive adaptability:

- Hypothesis 1: For a given task, individuals will exhibit different strategies with the specific strategy employed being a product of their intrinsic skills.
- Hypothesis 2: Individuals will exhibit varying levels of adaptability with an individual’s adaptability determining their propensity to switch strategies in response to changing circumstances.

To test these hypotheses, we are developing AKC techniques to allow us to characterize cognitive adaptability. Specifically, we will develop techniques to: (1) model patterns of selective information retrieval; (2) detect strategic biases revealing beliefs and intrinsic skills; (3) detect shifts in strategy over time; (4) develop mathematical techniques to bound the uncertainty in the individual cognitive models derived through AKC. We further intend to conduct experimental studies to establish neural correlates of behavioral metrics for cognitive adaptability.

3 Related Work

The study of individual differences has been a mainstay of psychology. Accordingly, a variety of traits, personality factors and performance dimensions have been discussed [2]. More recently, attention has focused on identifying neuro-physiological correlates of individual differences (e.g. Gevins & Smith, 2000). While psychological theories commonly accommodate individual differences and some fo-cus on explaining covariance in psychological measures across individuals, attention is generally focused on specific traits, as opposed to generalized mechanisms that ac-count for individual differences across a range of different dimensions. Furthermore, representations of cognitive theory within computational cognitive models have pro-vided provisions for adjusting various model parameters, but have offered little logic for adjustments beyond fitting the model to data obtained from a given experimental study [1; 5].

A central premise of the Cognitive Adaptability is that individuals differentially deploy their cognitive resources in response to ongoing circumstances. The same basic idea appears within other conceptualizations such as: Cognitive Continuum Theory (Dunwoody et al, 2000), which addresses judgments; Self-Organizing Cognition and dynamical systems approaches (Tschacher & Scheier, 1996; Tschacher & Dauwalder), which have been more heavily influenced by computer science than experimental cognitive research; and control theory applications to cognition (Jordan, 2000) which are based on engineering constructs that do not readily translate to biological systems.

4 Project Outline

Initial experiments will use a simple task in which subjects reproduce a line drawing within experimental conditions that place different demands upon their cognitive resources (e.g. retaining an image in working memory) or impose different task contingencies (e.g. different payoffs for speed vs. accuracy). Prior to experimental testing, separate measures will establish subjects proficiency for intrinsic skills associated with the experimental task (e.g. drawing precision, ability to handle mirror transformations) and personal biases (e.g. tendency to pursue high versus low risk rewards). Additionally, subjects' cognitive adaptability will be assessed using a response set switching paradigm (i.e. assessment of subjects' differential capacity to recognize that the rules governing a task have changed and adjust their behavior accordingly). It is hypothesized that for a given experimental condition, subjects will employ strategies that emphasize their individual cognitive strengths and biases. Furthermore, a subject's tendency to adopt strategies that emphasize skills for which they are less proficient or are contrary to their personal biases will vary in accordance with their cognitive adaptability. In the second year, the same paradigm will be employed but with a more complex task (the NASA Multi-Attribute Task Battery) that requires not only spatial skills, but also verbal processing, memory, and reasoning.

Current approaches for modeling cognitive task performance will be elaborated to encompass how an individual allocates their attention in performing a task. The resulting cognitive model will actively retrieve information from the task environment and exhibit information biases observed in the individual. To support automated knowledge capture, the task environment will be instrumented to include nonintrusive behavioral sensors such as eye tracking, posture recognition, mouse and keyboard manipulations, as well as a capacity to extract information from the graphical display including symbols, text, spatial positions and optical flow (i.e. movement of display elements in relation to one another).

5 Pilot Study

This section describes a pilot study of the line drawing task currently underway. The objectives of this study are 1) characterize strategies for the line drawing task, and 2) determine whether individual strategies correlate to aptitudes measured with a battery of standard psychometrics.

Fig. 1. A subject performing the line drawing task. Each subject in our pilot study draws for approximately 45 minutes.

5.1 Apparatus

The line drawing task is performed on a Wacom Cintiq 21UX interactive pen display. The display is approximately 43 cm wide by 33mm tall, however the drawing task is performed in a subregion approximately 22cm by 22cm. The resolution is 1200 rows and 1600 columns of pixels. This display was selected for its relatively large display/drawing area and low-latency response time (27ms claimed). All inputs are performed with the pen; the mouse and keyboard are not used by subjects in the experiment.

5.2 Drawing Task

Line drawing is performed in a software application shown in Fig. 2. The Picture Area (left) displays a figure to draw, while the Drawing Area (right) receives input. The picture need not be continuous and the subject may lift the pen and resume drawing at any time. The subject indicates completion of the trial by tapping outside the Drawing Area. Then the score for the trial is displayed briefly, then the task advances to the next trial. All of the settings described below are configured on a per-trial basis, so they can vary parametrically or randomly within a block of trials. An experiment session contains several blocks of trials. The order of blocks is randomized.

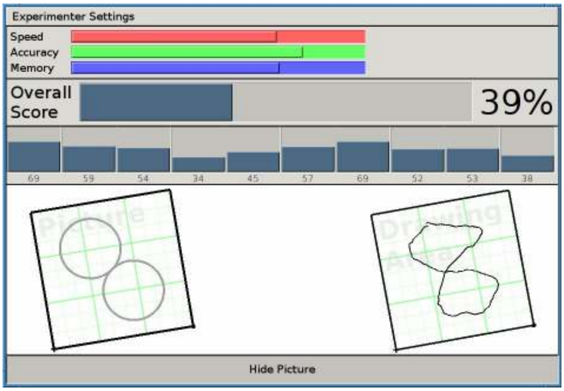


Fig. 2. The drawing task supports a variety of input and feedback conditions to elicit strategy shifts between and within subjects

5.3 Task Feedback

The task environment continuously scores each trial as the subject draws. There are three sub-scores and an overall score. Optionally, the scores are displayed and continuously updated to influence strategy selection.

- **Accuracy:** given sets of points $p_i \in P$ and $d_i \in D$ for the picture and drawing, respectively, with $0 \le p_{i_x}, p_{i_y} \le 1$ (and likewise for d_j) an accuracy score s_A is assigned according to Equation 1. The parameter τ determines how "strict" the metric is, with $\tau = 500$ a typical value. $\|p_i, D\|$ denotes $\min_{d_j \in D} \sqrt{(p_{i_x} - d_{j_x})^2 + (p_{i_y} - d_{j_y})^2}$ the distance from point p_i to the nearest point in set D, and $|P|$ is the number of points in set P.

$$s_A \equiv \left(1 - \frac{\sum_i \|p_i, D\|^2 + \sum_i \|d_i, P\|^2}{|D| + |P|}\right)^{\tau} \tag{1}$$

- **Speed:** the speed score s_S is a decay function of t, the duration of the trial, with parameter $t_{1/2}$ specifying the number of seconds before the score decays to 0.5 (Equation 2):

$$s_S \equiv \left(\frac{1}{2}\right)^{\frac{t}{t_{1/2}}} \tag{2}$$

- **Memory:** the memory score s_M is also defined by Equation 2, but taking the place of t is t_v, the number of seconds the picture has been visible during the current trial. Thus the memory score is maximized by viewing the picture only briefly. The memory score is calculated only in trials where the subject manually shows and hides the picture. There is a forced delay (typical value 2s) after each time the picture shown, which imposes a fixed penalty for each viewing (through the speed score) and requires the subject to hold the picture in memory. The picture is initially hidden, so the subject may predict the next picture in the sequence of trials to achieve a perfect memory score, at the cost of a low accuracy score if the prediction is incorrect.
- **Overall:** the overall score s_O combines the subscores $S = \{s_A, s_S, s_M\}$, each with a corresponding weight $0 \le \alpha_i \le 1$ in Equation 3. Thus renders the corresponding metric entirely moot, while $\alpha_i = 1$ implies that the overall score cannot be higher than the subscore.

$$s_O \equiv \prod_i 1 - \alpha_i + \alpha_i S_i \tag{3}$$

If required by the experiment design, visual feedback is presented by displaying the composite and overall scores graphically and numerically (**Fig. 2**). Below the overall score is a graph which shows proceeding trial scores, which may help a subject identify performance trends and motivate him or her to improve over time. The score display is updated at 10hz.

Significantly, the user interface does not display the system parameters (e.g. speed score half-life, nor subscore weights). For a good score, a subject must develop a strategy that is both consistent with their abilities, and which is rewarded by the environment at the time.

5.4 Task Manipulations

The drawing software supports several task manipulations to elicit strategy shift between and within subjects, including:

- **Picture:** The picture being drawn may be familiar or novel, detailed or simple, sharp-cornered vs. smooth, etc.
- **Affine Transformation:** The position, scale, and orientation of the Picture and Drawing areas can be set independently, forcing the subject to mentally transform the picture.
- **Drawing vs. Tracing:** The Picture and Drawing areas may coincide, resulting in a tracing task.
- **Memory:** The picture may be hidden and a delay imposed before drawing, forcing the subject to draw from memory.
- **Timeout:** Drawing time may be limited. The timeout is a normally distributed random variable invisible to the subject. This condition prompts the subject to choose between reliably earning a lower score by drawing quickly, or drawing more

slowly in hopes of a higher score at the risk of receiving no credit if the timeout is exceeded.
- **Interstimulus Interval:** The delay between trials is varied.
- **Background:** The backgrounds displayed in the Picture Area and Drawing Area (e.g. a grid) can be used to vary landmarks for the drawing task.
- **Invisible Drawing:** The marks drawn by the pen may be hidden (as if the pen were out of ink). This requires the subject to remember which parts of a figure have been completed and makes it harder to identify errors, decreasing the accuracy score.

5.5 Output

For each trial, the drawing task software outputs the following information:

- Each point in the picture
- Each point drawn by the subject, with time stamps. The sample rate averages 140 Hz which is limited by the windowing system (X.Org X Server 1.5.2 on Ubuntu Linux 8.10).
- All of the settings in effect during the trial
- The duration of the trial and the scores displayed to the subject.

6 Conclusion

This paper proposed that cognitive adaptability is a trait necessary to explain the inherently dynamic nature of cognitive processes as individuals adapt their available resources to ongoing circumstances. We outlined a research plan that is intended to establish cognitive adaptability by measurement and prediction of behavioral data and the discovery of neural correlates. Subsequent papers will document experiments and findings from this course of research.

References

1. Anderson, J.R., Lebiere, C.: The atomic components of thought. Erlbaum, Mahwah (1998)
2. Cooper, C.: Individual Differences. Arnold, Belfast (2002)
3. Miller, M.B., Van Horn, J.D., Wolford, G.L., Handy, T.C., Valsangkar-Smyth, M., Inati, S., Grafton, S., Gazzaniga, M.S.: Extensive individual differences in brain activations associated with episodic retrieval are reliable over time. Journal of Cognitive Neuroscience 148, 1200–1214 (2002)
4. Taylor, S.F., Welsh, R.C., Wager, T.D., Luan Phan, K., Fitzgerald, K.D., Gehring, W.J.: A functional neuroimaging study of motivation and executive function. NeuroImage 21(3), 1045–1054 (2004)
5. Wray, R.E., Laird, J.E.: An architectural approach to consistency in hierarchical execution. Journal of Artificial Intelligence Research 19, 355–398 (2003)

Implications of User Anxiety in the Evaluation of Deception in Web Sites

Brent Auernheimer[1], Marie Iding[2], and Martha E. Crosby[3]

[1] Digital Campus, California State University, Fresno, CA 93740, U.S.A.
[2] University of Hawaii, College of Education, 1776 University Avenue, Honolulu, HI 96822, U.S.A.
[3] University of Hawaii, Department of Information and Computer Sciences, Post 317A, Honolulu, HI 96822, U.S.A.
{auernhei,miding,crosby}@hawaii.edu

Abstract. In a pilot study, we investigated the effect of anxiety on users' susceptibility to deceptive information on Web pages. Specifically, we manipulated the perceived control and associated anxiety of participants with and without visual disabilities as they used an assistive technology, a screen reader. Preliminary findings indicated that anxious participants (i.e., without visual disabilities) using the unfamiliar assistive technology were more susceptible to deception *and* expressed more suspicion regarding the Web pages. We interpret these preliminary findings as consistent with the work of Whitson and Galinsky [1] and discuss implications for further research in Web site credibility determinations and users' susceptibility to deception.

Keywords: Deception, Web Sites, Anxiety, Control, Assistive Technology.

1 Introduction

Why do people fall prey to deceptive information on the Web? What factors affect their susceptibility to deception? While it is clear that the information itself and how it is presented effects people's determinations of the credibility of Web-based information [2, 3], it is not yet clear to what extent user characteristics or user contexts can affect those determinations. For example, are users more or less vulnerable to deception under anxiety-provoking circumstances such as when they experience a decreased sense of control? In this paper, we examine the use of an assistive technology for reading and its effects upon the anxiety levels of persons with and without visual impairments as they assess deception in Web pages. In initiating this work, we draw upon several research areas from which we briefly describe relevant findings: characteristics of users that are known to affect credibility determinations, research examining illusion of control and people's perceptions of distortions [1], and basic educational work with assistive technology.

1.1 Characteristics of Users

Several characteristics affect user determinations of credibility including socioeconomic status, educational and reading levels. For example, Benotsch et al. [4]

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 114–119, 2009.

compared HIV positive patients' and medical professionals' ratings of a credible Web site presenting HIV treatments (from *Journal of the American Medical Association* or *JAMA*) to that of a Web site describing a purported cure for HIV involving goat serum. Medical professionals and those from higher socioeconomic, educational and literacy levels were better able to discriminate the difference in information quality than those from lower SES and literacy levels. Benotsch et al argue that vulnerable populations need to receive special instruction in Web site/information evaluations so as to be less vulnerable to deceptive cures and misinformation.

In work by Iding et al. [5] more highly educated participants were more confident about their determinations of Web site credibility in a very specific area of software engineering. Less educated (e.g., students) were less confident, but in this particular case, were equally accurate in their credibility determinations. While this result would seem incontrovertible, in other research, among film studies students in Norway, Iding, et al. [5] found that the more educated people were, the less confident they were about their credibility determinations. Upon initial consideration, these findings appear counterintuitive. Although without further research we cannot fully account for these differences, it appears that the software engineering topic that was the focus of the former research was a narrow topic that had been covered in classes and was known to participants. The film studies Web sites, although selected by professors in the field, covered films and related topics that were had a lesser probability of being specifically known to participants. We speculate that this field might be more diffuse than a specific software engineering topic and so might account for people's greater awareness of their knowledge limitations with education.

These are examples of only a few user characteristics that can be associated with people's abilities to adequately assess credibility of Web pages and detect deception. How do contextual factors affect people's determinations?

1.2 Sense of Control and Pattern Perception

In "Lacking Control Increases Illusory Pattern Perception," Whitson and Galinsky [1] describe research in which people's sense of control is reduced in various conditions. Participants tended to perceive patterns in arrays of stimuli where none actually exist. Whitson and Galinsky explain: "Participants who lacked control were more likely to perceive a variety of illusory patterns, including seeing images in noise, forming illusory correlations in stock market information, perceiving conspiracies and developing superstitions…. when individuals are unable to gain a sense of control objectively, they will try to gain it perceptually" (p. 115) [1]. This is confirmed by Rudski's research with college students who tended to rely on rituals and superstitions when they were less prepared, less in control of the outcome, and the stakes were high [7].

What effect would manipulating people's sense of control and related anxiety have upon their ability to detect deception in Web pages? In the present research, we hypothesized that reducing control could be achieved for participants without visual disabilities by using an unfamiliar assistive technology, a screen reader, for examining Web pages. We would expect these participants to be less accurate in detecting deception and more apt to perceive conspiracies in accurate Web-based information. Participants with visual disabilities would be familiar with the assistive technology, so we hypothesized that the opposite would happen: they would experience a heightened sense of control and reduced anxiety and thus be less susceptible to deception on Web pages.

1.3 Assistive Technology for Reading on the Web

Because the present research utilizes assistive technology as a means to manipulate users' "sense of control" and associated anxiety, an introduction to assistive technology for reading and its relationship to user control merits discussion. Anderson-Inman and Horney [8] describe "text that has been altered to increase access and provide support to learners as *supported electronic text* or *supported e-text*" (p. 153), and provide a brief overview of kinds of supports possible, "Including embedded supports (e.g., definitions of unfamiliar terms), multiple modalities (e.g., text that can be read out loud), and links to useful resources (e.g., background information, concept map, notepad) – all of which can transform electronic text so that it is more accessible and supportive to diverse learners" (p. 153). The authors also describe a typology for supported eText adapted by the National Center for Supported eText (NCSeT).

They mention that a case cannot be made that text-to-speech devices when used alone improve comprehension for users with visual disabilities. For example, they describe prior work in which users with hearing disabilities preferred to use a pronunciation tool although it had been deemed not useful for this group. The authors contend, "Personal choice when interacting with supportive resources can be highly motivating in itself, resulting in increased engagement with the text and potential for increased comprehension" (p. 158).

This conclusion is corroborated by the work of Badge et al. [9], who found that a group with disabilities used far more features for creating PowerPoint presentations for the Web than their counterparts without disabilities. As they explain, "It is possible that these students [with disabilities] were used to customizing their own learning experiences and personalizing their computing environment and as such were perhaps more self-aware than the control group who mostly appeared to passively watch the presentations with little interaction" (p. 111).

2 Method

In this preliminary study we examine the relationship between users perceived sense of control and how they make determinations of deception in websites. Our research question was: How does user control affect credibility determination? Our hypotheses were: 1) that in environments where the participants felt that they lacked control, they would be more likely to consider the websites to be deceptive and 2) when participants were in an environment where they felt in control, they were more likely to give accurate assessments of the websites veracity. Control was measured in two ways.

First, the environment was manipulated by having participants view some Web pages, and only hear the information that existed on other Web pages. Second, the results of the Rotter's Locus of Control Inventory should indicate the extent participants consider themselves generally "in control." In addition to their credibility judgments, we were also interested in how much control that the participants felt that they had, as well as their confidence in their own credibility judgments.

2.1 Participants

Participants consisted of one person with a visual disability and 13 individuals without visually disabilities recruited from the University of Hawaii and California State University, Fresno.

2.2 Materials and Procedure

Before the participants without a visual disability begin the experiment, they installed the Fire Vox add-on to the Firefox browser, and make sure that they could start the "automatic read mode" to read entire Web pages. Fire Vox is an open source, multi-platform screen reader, freely available from http://firevox.clcworld.net. The Fire Vox developers describe it as:

> A screen reader for everyone - Fire Vox is designed to accommodate different users with different needs. For visually impaired users, all Fire Vox commands are keyboard activated. In addition, the keyboard commands can be easily reconfigured in the self-voicing Fire Vox Options menu to avoid conflicts with other accessibility software products or to suit personal preferences. For sighted users who need a screen reader, such as web developers interested in testing their webpages or educators who work with visually impaired students, Fire Vox's highlighting feature makes it easy to keep track of where it is reading from on a page. This highlighting feature is also useful for dyslexic users and partially sighted users.

The participant with a visual disability used the JAWS commercial screen reader (http://www.freedomscientific.com/) that they was accustomed to. For the other participants the Fire Vox screen reader was an unfamiliar and uncomfortable environment.

We categorized three of the websites as accurate and the other three as deceptive. We found the six Web pages using personal knowledge and search engines. The deceptive Web pages were either purposely deceptive (such as http://www.dhmo.org/) or described ideas not recognized by mainstream science. The accurate Web pages described real, but unusual, information or situations.

Next, an order for presentation of the Web pages was randomly determined. Then within that order, Web pages were assigned at random to be seen or heard. To balance the experiment a second list of the same Web pages was created in the same order as the first, but with "seen" and "heard" flipped. So, participants without a visual disability heard half the Web pages and saw half the Web pages. The participant with a visual disability participated in the "sound only" interface where they felt comfortable and in-control.

The participants visited six Web sites and in turn completed a survey about each page. The survey asked them whether or not they believed the information on the Web page and how confident they were in their answer. The questionnaires gave the participants a choice of seven rating levels.

In addition to the questionnaires about the Web pages, the participants answered the following follow-up questions at the end of the study:

- "In general, rate the level of control that you felt while carrying out this task"
- "Rate the level of anxiety that you felt while carrying out this task"
- "Please add any comments that you might have about your experience completing these tasks."

Participants then completed Rotter's twenty-nine item Locus of Control Inventory [10]. The inventory measures the extent to which people attribute outcomes to their own efforts or external factors.

3 Results and Discussion

The authors coded the participants' written comments in response to each question in terms of a range of possibilities from accurate to deceptive. Even with the limited data set, the data suggests that if the participants, regardless of whether they had a visual disability, were comfortable and felt in control, they were more likely to discern which Web pages were accurate and which ones were deceptive. On the other hand, if the participants were forced into an unfamiliar environment where they were uncomfortable and felt as though they little control, they were unable to discriminate the accurate from the deceptive Web pages as well. Interestingly, the Rotter's Locus of Control Inventory did not seem to have as much influence on the participants' decisions as the authors anticipated. Findings from this pilot experiment highlighted contradictions between what some participants viewed as credible and others did not. Furthermore, the results pointed out differences in confidence about one's ability to evaluate information on the Web in general.

Finally, we note that all users can find assistive technologies helpful. For example, consider a user-submitted "strategy" on the Lexdis website (http://www.lexdis.org/strategy/152) helpful to users with and without a visual disability:

> The first thing I had to do was use text-reading software over some poetry, just to slow myself down. Text-reading software is atonal, finding recordings of people reading is a much better option, but in this case it had the desired effect of making me pay attention to every word.

The approach of "universal usability" [11], an extension of the physical Universal Design movement (http://www.design.ncsu.edu/cud/), is a rich research area for Web and interface designers. As richer interfaces become widespread, users will be more comfortable with technologies such as screen readers -- much as we've become accustomed to sidewalk curb cuts and ramps – and all will benefit. Sarah Horton (http://universalusability.com/) quotes Ben Shneiderman's definition of universal usability: "enabling all citizens to succeed using communication and information technology in their tasks".

References

1. Whitson, J.A., Galinsky, A.D.: Lacking Control Increases Illusory Pattern Perception. Science 322, 115–117 (2008)
2. Iding, M., Auernheimer, B., Crosby, M.E., Klemm, E.B.: Web Site Credibility: Why do People Believe What They Believe? Instr. Sci. 37, 43–63 (2009)
3. Fogg, B.J., Soohoo, C., Danielson, D., Marable, L., Stanford, J., Tauber, E.R.: How Do People Evaluate a Web Site's Credibility: Results from a large study (2002), `http://www.consumerwebwatch.org/dynamic/web-credibility-reports-evaluate-abstract.cfm`
4. Benotsch, E.G., Kalichman, S., Weinhardt, L.S.: HIV-AIDS Patients' Evaluation of Helath Information on the Internet: The Digital Divide and Vulnerability to fraudulent claims. J. of Consulting and Clin. Psych. 72, 1004–1011 (2004)
5. Iding, M., Auernheimer, B., Crosby, M.: Towards a Metacognitive Approach to Credibility. In: Proceedings of the Second Annual Workshop on Information Credibility on the Web in Conjunction with ACM 17th Conference on Information and Knowledge Management, CD ROM - Workshop Proceedings, pp. 76–80 (2008)
6. Iding, M., Nordbotten, J., Singh, J.M.: Credibility: Norwegian Students Evaluate Media Studies Web Sites. In: Kuma, D., Turner, J. (eds.) Education for the 21st century: Impact of ICT and digital resources: WCC 2006, Santiago, Chile, pp. 327–331. Springer, New York (2006)
7. Rudski, J.M.: Malinowski Goes to College: Factors Influencing Students' Use of Ritual and Superstition. J. of Gen. Psych. 134, 389–403 (2007)
8. Anderson-Inman, L., Horney, M.A.: Assistive Technology Through Text Transformations. Reading Res. Quarterly 42, 153–160 (2007)
9. Badge, J.L., Dawson, E., Cann, A.J., Scott, J.: Assessing the Accessibility of Online Learning. Innovations in Ed. And Teaching International 45, 103–113 (2008)
10. Rotter, J.B.: Generalized Expectancies for Internal, External Control of Reinforcement. Psychological Monographs 80, 1–28 (1966)
11. Shneiderman, B.: Universal usability. Communications of the ACM 43(5), 84–91 (2000)

Investigation of Sleepiness Induced by Insomnia Medication Treatment and Sleep Deprivation

Ioanna Chouvarda[1], Emmanouil Michail[1], Athina Kokonozi[1],
Luc Staner[2], Nathalie Domis[2], and Nicos Maglaveras[1]

[1] Lab of Medical Informatics, the Medical School, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece
[2] FORENAP, Rouffach, France
{nicmag,ioanna,michem,athko}@med.auth.gr,
{Luc.STANER,Nathalie.DOMIS}@forenap.com

Abstract. The main objective of this work is the study of EEG signals in order to investigate sleepiness induced from drug administration for insomnia and sleep deprivation. Data used in this work were obtained from real experiments in FORENAP, France and in CERTH, Thessaloniki, Greece. The features under consideration are Power Spectrum in certain frequency areas, alpha slow-wave index (ASI) and Fractal Dimension (FD) for placebo and verum subjects. Studying these features in the above groups, we found that sleepiness due to hypnotic medication and due to sleep deprivation can cause different behaviour in brain activity at certain locations. These EEG characteristics could be used for the classification of the medication intake (verum or placebo) and its effect.

Keywords: insomnia, sleep deprivation, EEG signals.

1 Introduction

Insomnia is a medical disorder of sleep patterns characterized by difficulty in falling asleep, remaining asleep, or both. It affects millions of people and can be caused by many different conditions, diseases, and circumstances. Some effective insomnia treatments focus on changing the sleep behaviours and habits, while others require medications and supplements.

It is known that in humans sleep does not begin the same time in all cortical areas. Topographical and frequency changes observed in EEG data effect the wakefulness-sleep transition and allow us to describe the state of human brain before and after sleep on set [1].

Sleep deprivation is an overall lack of the necessary amount of sleep. Changes in brain activity have been observed during sustained wakefulness. Assessment of EEG power density in sleep deprived people [2] demonstrated an incensement in the 6.25-9.00 frequency range. Fluctuations in the energy of theta and alpha bands can be electrophysiological correlated to the “waking intensity” [3], [4]. Sleep-deprived subjects showed shifted patterns of brain activity, but research in this area is still controversial.

In this work we examine the coexistence of EEG topographical and frequency changes in order to elucidate differences in sleepiness due to hypnotic drug administration and due to sleep deprivation.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 120–127, 2009.

2 Methods

In this analysis two datasets were used: a) dataset with drug induced sleepiness, b) dataset with sleepiness induced by sleep deprivation.

Concerning the first one, 14 male subjects aged from 18 to 40 years were selected as volunteers, in FORENAP. EEG data were collected after lorazepam 2.5 mg single administration in the morning to healthy people participated in the study. Lorazepam is a benzodiazepine drug with short to medium duration of action. It is known for its anxiolytic, amnesic, sedative/hypnotic, anticonvulsant and muscle relaxant properties by slowing down the central nervous system. As a psychoactive drug, it is useful in treating insomnia. For each subject two conditions were examined: verum and placebo, with recordings corresponding to different times during the day and night until next morning (recordings: t_1: one hour before drug intake, t_2: drug intake, t_3:one hour after drug intake, t_4:two hours after drug intake, t_5: three hours after drug intake, t_6:four hours after drug intake, t_7:five hours after drug intake, t_8:six hours after drug intake, t_9:eight hours after drug intake, t_{10}:ten hours after drug intake, t_{11}:twelve hours after drug intake, t_{12}:thirteen hours after drug intake). EEG signals were obtained with eyes closed during 3 minutes in resting condition. Standard channels used : chan.1 : FP1, chan.2 : FP2, chan.3 : F7, chan.4 : F3, chan.5 : FZ, chan.6 : F4, chan.7 : F8, chan.8 : T3, chan.9 : C3, chan.10 : CZ, chan.11 : C4, chan.12 : T4, chan.13 : T5, chan.14 : P3, chan.15 : PZ, chan.16 : P4, chan.17 : T6, chan.18 : O1, chan.19 : 0Z, chan.20 : O2.

For the pre-processing of the data, the average of all channels were calculated as reference and subtracted from all channels. Designated artifact regions were zeroed. Basic filtering was done at: 0.5-25 Hz. During the basic processing part, the standard 20 channels mentioned before have been used for analysis, not the extra ones.

The second dataset was obtained from an experiment [5] that took place at CERTH, Thessaloniki, Greece, from 6 June till 27 July 2005. Subjects participated in this one, were average drivers (mean driving experience: 8.3 years), with a mean 26.5 years and were asked to stay awake for at least 24 hours. The level of sleepiness was estimated by using the Karolinska Sleepiness Scale (KSS), [6] ranging from 1(very alert) to 9(very sleepy). The KSS test has been found to be related to EEG and behavioral variables, indicating a high validity in measuring sleepiness [7].

Data acquisition was performed for 20 minutes in a quite, dark environment. Recordings corresponding to the last 3 minutes were analyzed, in correlation with the data from the previous experiment. A sampling rate of 200Hz was used and the amplitude range was ± 20μV. Band bass filtering at the range of 0.5 to 70Hz was applied, with a notch filter ate the 50Hz power supply component.

During preprocessing [8], EEG data were Band pass filtered (3rd order Butterworth filter, Band pass range: 0.5 – 45 Hz) and artifacts were removed by Independent Component Analysis (ICA) technique. Finally, data were filtered to 0.5-25Hz, in order to have the same spectral range as the first dataset.

For the assessment of differences in brain activity between subjects manifesting sleepiness under insomnia medication (verum and placebo groups) and sleep deprived subjects with manifested sleepiness, spectrum analysis was applied. The extracted features were Power Spectrum in the frequency bands: delta (1-4Hz), theta (4-8Hz), alpha (8-13 Hz), beta (13-22 Hz) and ASI- alpha slow-wave index

(ASI=alpha/(delta+theta)) , related to arousal level. Fractal Dimension (FD) was also calculated for EEG signals, an indicatr of the system's complexity, related to both arousal/sleepiness and vigilance.

3 Results

Using the aforementioned features, the verum/placebo differences before medication were first assessed, forming the baseline for the drug effect. Then the differences between the two conditions were assessed again for the recordings one hour after medication intake, when drug effect is expected to be high. In parallel, these characteristics were compared with the ones of sleep- deprived subjects manifesting sleepiness.

Power spectral analysis revealed important differences between verum and placebo group in delta, alpha and beta frequency ranges for certain scalp locations.

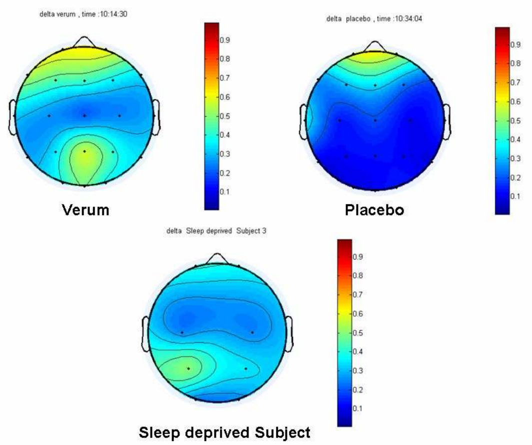


Fig. 1. Topographic map of a scalp data field with specified channel locations, show the brain activity for a subject for delta band, one hour after lorazepam administration and for a sleep deprived subject

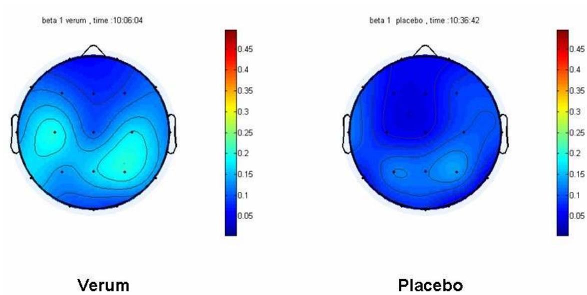


Fig. 2. Energy distribution for a verum and placebo subject one hour after drug intake at 13-22Hz

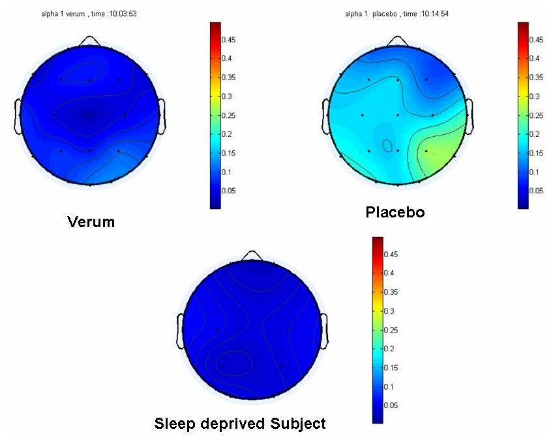


Fig. 3. Energy distribution for a verum and placebo subject one hour after drug intake and for a sleep deprived subject, for the alpha band

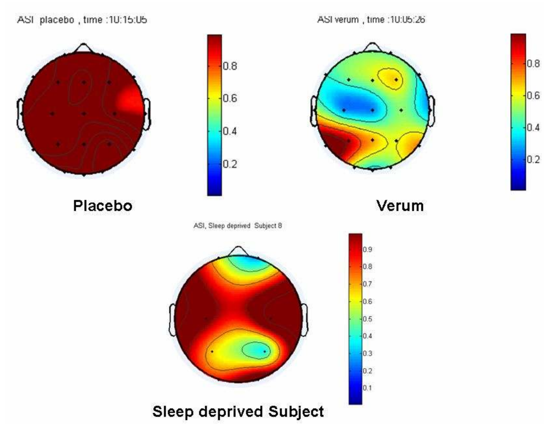


Fig. 4. Topographic map of a scalp data field with specified channel locations, show ASI one hour after lorazepam administration for a subject under verum - placebo conditions and for a sleep deprived subject

More specifically, for 1-4Hz frequency range and for the verum group, after drug intake, a predominance was observed for channels Fp1, F7, F3, Fz, P3, P4, and Pz. Placebo subjects showed lower energy values in comparison with verum and sleep deprived subjects. A topographic map of a scalp data field in a 2-D circular view shown in Fig.1, illustrates energy distribution at 1-4Hz, for a verum and a placebo subject one hour after drug administration and for a sleep deprived subject.

Energy increases for channels F3, C3, P3, P4, Cz, C4, T5 and T1 in beta (13-22 Hz) band for subjects that took lorazepam, one hour after the drug intake. Fig.2 shows an example with the energy distribution in the channels mentioned before for a verum and placebo subject, one hour after having 2.5 mg of lorazepam.

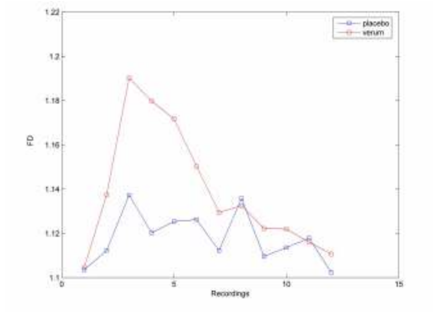


Fig. 5. Evolution of FD with time, for a specific subject and channel C3

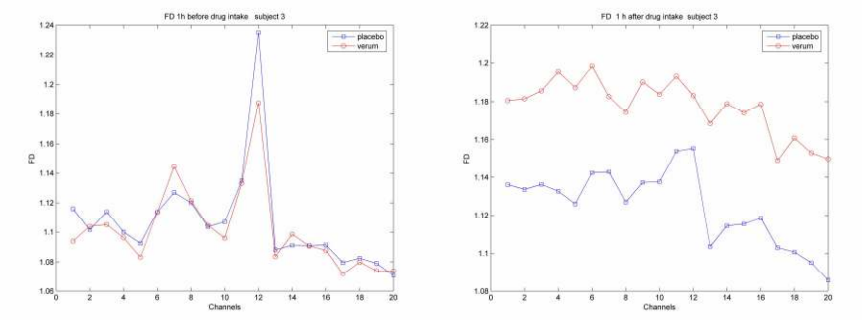


Fig. 6. FD for all channels, one hour before drug intake and FD for all channels, one hour after drug intake

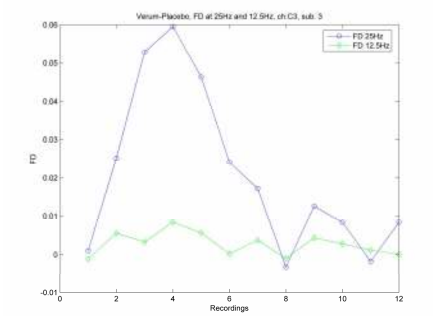


Fig. 7. FD difference between verum and placebo (FDv - FDp) for filtered signals at 25 Hz and 12.5 Hz for one subject

After one hour of drug administration, energy for alpha band (channels: F3, C3, Cz, C4, T5, T6) decreased for verum subjects, compared to placebo, and the alpha band levels where lower than in sleep-deprived subjects (Fig.3).

ASI feature indicating arousal decreases with verum in comparison not only with placebo, but also with sleep deprived group, since energy in delta and theta frequency

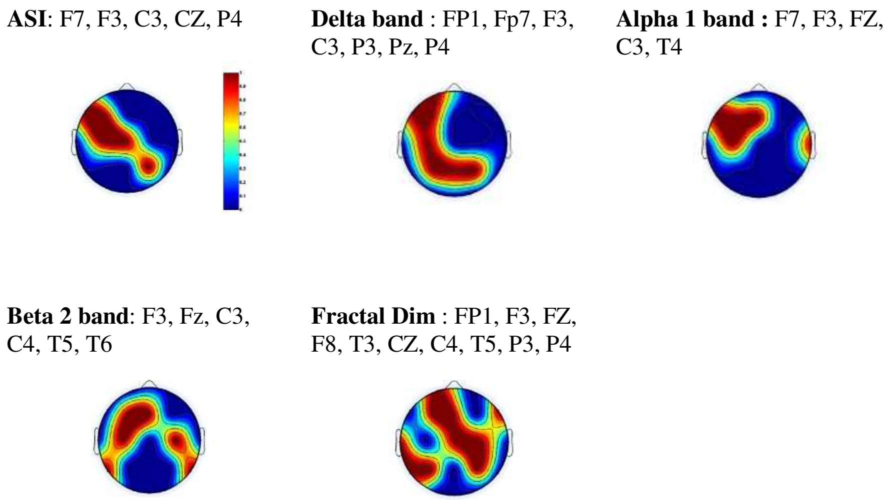


Fig. 8. Topographic map of a scalp data field with specified channel locations, show with red the positions that correspond to the statistically significant changes described above between verum and placebo, one hour after drug intake

bands for subjects under medication has higher values in comparison with sleep deprived subjects. An example can be seen in Fig.5.

EEG activity across delta/theta/sigma (12-15Hz) frequency range for channels FP1, FP2, F7, F3, Fz, F4, F8, C3, Cz, C4 was also significant different between verum and placebo groups, for all the EEG recordings.

Finally FD feature was calculated for subjects that took hypnotic drug and for placebo subjects. FD for verum group appeared significantly higher than in placebo group, while these differences dimish after 24 hrs. Fig.5 shows the FD evolution for C3 channel for a specific subject. In Fig.6 FD is plotted versus channels for the same subject, one hour before drug intake, while Fig.7 shows the situation one hour after drug intake, both for verum and placebo subjects. The difference between FD values for verum and placebo decreases by filtering the EEG signals, as shown in Fig.9 at 25 Hz and 12.5 Hz, indicating that the observed increased energy in beta band for verum could be partly responsible for the altered FD values in this group. However, more studies are needed, as fractal dimension is known to be affected by the EEG signal bandwidth. Furthermore, it is characteristic that on average the fractal dimension of the sleep deprived subjects who manifested sleepiness was also high, in comparison with the FD measured in the other subject group.

Overall, statistical differences were detected with the Wilcoxon rank sum test. Comparing the verum/placebo features before drug intake, statistical differences were not found in any EEG channel. On the other hand, one hour after drug intake, statistically significant changes were found between verum and placebo in ASI, Fractal Dimension, and bands Delta, Alpha1 ans Beta 2, as depicted in Fig.8 and Table. 1.

Table 1. P-values for Wilcoxon rank sum test results, corresponding to differences in EEG channel recordings as described above, between verum and placebo subjects, one hour after drug intake

	ASI	Delta	Alpha 1	Beta 2	FD
FP1		0.0409			0.0229
F3	0.0366	0.012	0.0229	0.0203	0.0409
F5			0.0409		
FZ				0.0291	0.0456
F8					0.0366
F7	0.0123		0.0409		
C3	0.0258	0.0203	0.0229	0.0456	
C4				0/0366	0.0456
CZ	0.0159				0.0291
P3		0.0229			0.014
P4	0.0326	0.0054			0.0336
PZ		0.0366			
T3					0.0291
T4			0.0159		
T5				0.0366	0.0203
T6				0.014	

4 Discussion

The analysis in this work demonstrates that power spectrum energy and ASI for certain EEG channels can be characteristics different for the three subject groups under consideration. Studying these features in the above groups, we found how sleepiness after drug intake is reflected in the EEG features, and moreover that sleepiness due to medication (one hour after drug intake) and due to sleep deprivation can cause different behaviour in brain activity at certain locations. Drug intake causes increase in delta/beta band and decrease in alpha band, as well as an increase of the fractal dimension in almost all channels.. Differences were not significant in the occipital channels, but rather in the left centro-parietal area. Furthermore, it is interesting to note that ASI index as well as fractal dimension was higher for the sleep deprived than medication group, suggesting that drowsiness or sleepiness due to medication is higher, or that the sleep-deprived group maintains more mental ability than the medication group. However, the fact that data for these two groups (medication and sleep deprivation) were produced from two distinct experiments consists a limitation to this study. Furthermore, extended experiments would be required to reveal to what extent these conditions cause a combination of sleepiness and hypovigilance, or preferably one of them.

Concluding, altered FD of EEG signals could support the detection of brain patterns in verum group, for the specific drug administration. Spectral characteristics discussed in this paper could also address the detection of drug administration effects, discriminating between verum and placebo, and also between medically and naturally induced sleepiness.

Acknowledgments

This work has been partly funded by the EU-IST–2003 - 507231 project SENSATION.

References

1. De Gennaro, L., Ferrara, M., Curcio, G., Cristiani, R.: Antero-posterior EEG changes during the wakefulness-sleep transition. Clinical Neurophysiology 112, 1901–1911 (2001)
2. Cajochen, C., Brunner, D.P., Kräuchi, K., Graw, P., Wirz-Jstice, A.: Power density in theta/alpha frequencies of the waking EEG progressively increases during sustained wakefulness. Sleep 18, 890–894 (1995)
3. Åkerstedt, T., Gillberg, M.: Subjective and objective sleepiness in the active individual. Int. J. Neurosci. 52, 29–37 (1990)
4. Makeig, S., Jung, T.P.: Changes in alertness are a principal component of variance in the EEG spectrum. Neuroreport 7, 213–216 (1995)
5. Portouli, E., Bekiaris, E., Papakostopoulos, V., Maglaveras, N.: On-road experiment for collecting driving behavioural data of sleepy drivers. Somnologie 11(4), 259–267 (2006)
6. Åkerstedt, T., Gillberg, M.: Subjective and objective sleepiness in the active individual. Int. J. Neurosci. 52, 29–37 (1990)
7. Kaida, K., Takahashi, M., Åkerstedt, T., Nakata, A., Otsuka, Y., Haratani, T., Fukasawa, K.: Validation of the Karolinska sleepiness scale against performance and EEG variables. Clinical Neurophysiology 117, 1574–1581 (2006)
8. Papadelis, C., Chen, Z., Kourtidou-Papadeli, C., Bamidis, P., Chouvarda, I., Bekiaris, E., Maglaveras, N.: Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accident. Clinical Neurophysiology 118, 1906–1922

Activity Awareness and Social Sensemaking 2.0: Design of a Task Force Workspace

Gregorio Convertino, Lichan Hong, Les Nelson, Peter Pirolli, and Ed H. Chi

Palo Alto Research Center (PARC)
3333 Coyote Hill Road
Palo Alto, CA 94304, USA
{convertino,hong,lnelson,pirolli,echi}@parc.com

Abstract. Task forces of expert knowledge workers would benefit from more advanced web tools supporting activity awareness and social sensemaking. This paper proposes the design of a task force workspace, which is under development. It introduces the problem through a scenario, specifies requirements, illustrates a modeling approach and the mockups of the functions in the proposed workspace. Design issues and future work are finally discussed.

Keywords: Awareness, Sensemaking, Task Force, Roles, User Modeling, CSCW Design, RSS or Atom Feeds.

1 Problem

Numerous and diverse task forces of expert knowledge workers would benefit from more advanced web tools supporting awareness and social sensemaking. Examples of task forces are a national scientific committee writing an official report on climate change, a team of professionals writing a proposal for a large contract bid on behalf of their company, a corporate group learning about and identifying opportunities in a new business area, a group of financial analysts or a tiger military team strategizing a solution to a complex problem, an inter-agency task force planning and managing responses to a major natural disaster such as a hurricane.

The tasks of these large work groups tend to be novel in topic, ad hoc in method, and have a number of constraints in time and space (e.g., limited time, asynchronous work, and distributed across different places). The large amount of labor, the broad variety of skills required, and the critical implication of the final product make it necessary that multiple experts contribute to the work. This paper describes an ongoing research that focuses on supporting a specific set of needs of these task forces. Such needs include managing a large amount of noisy data, summarizing data coming from multiple sources, and coordinating among collaborators with diverse roles toward the common goal of delivering a final report (or a solution to a problem) which synthesizes the information content foraged and the judgments made on it.

Two classes of tools help task forces, respectively, to collect information and author a final report. Web tools have expanded knowledge workers' abilities in foraging

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 128–137, 2009.

large numbers of tokens or chunks of information, such as web pages, wiki pages, blogs, documents (or portions of them) and share them with others (e.g., through email, search engines, feed readers or aggregators, shared databases). Moreover, workers typically have access to collaborative editing applications for collaboratively writing reports (e.g., wikis, Google Docs, groupware applications). Despite the abundance of tools in these two classes, currently there are very few tools to assist the workers in doing the in-between work of filtering, abstracting, and organizing low-level tokens of information into intermediate representations that progress towards the components of the desired final product (e.g., see preliminary attempts in 10, 10]. That is, in a nutshell, we lack social sensemaking tools.

This paper proposes a workspace design that supports awareness, monitoring, and social sensemaking in a task force. The next sections present a scenario and, in relation to it, illustrate requirements for task forces. Then, we describe the functions of our prototype and briefly discuss the main design issues and future research.

2 Context

Task scenario

Let us consider a real problem scenario for a task force.

The US government establishes a scientific task force on climate change. The task force includes about thirty-five members with very diverse specialties: biologists, economists, climatologists, lawyers, policy analysts, and other professionals. The goal is to identify the science and information needed to assist the government in addressing the consequences of climate change and to suggest possible options for getting the needed science. The concrete task is to produce a progress report in a period of about 6-9 months. The task force needs to forage and summarize large amount of information from various digital sources such as scientific libraries, government databases, the Internet, personal media (e.g., email, private databases). The members need to share and discuss the relevant chunks of information, then write and assemble sub-sections of the progress report in a shared wiki. That is, they generate intermediate summaries that are later used to compose the final proposal.

This problem scenario is modeled after a real task force formed in 2007 to identify the science and information needed to assist the government in addressing the consequences of climate change and to suggest possible options for doing the needed science [1]. It exemplifies aspects of a task force that are useful when specifying design requirements, which are:

1. *Specialized co-workers*: large group of knowledge workers, including a chair and domain experts of very diverse backgrounds as members.
2. *Collaborative task*: progress report writing is a complex knowledge task requiring labor division across experts who share a goal, i.e., delivering a high-quality report.
3. *Setting*: the work is distributed across places, the projects is completed over several months, members collaborate mostly asynchronously with few coordination meetings.
4. *Tools*: Web, various databases, and a wiki for drafting the progress report.

Design Goal

The goal is to engineer tools as part of a web-based workspace that ultimately improve the quality of the task force's final report by *reducing the costs (or increasing the benefits) for the members to*:

- Construct and share intermediate sensemaking products.
- Maintain awareness of relevant content and roles of contributors.

The theory guiding the design is the sensemaking model [2] and research on collaborative tools supporting awareness and sensemaking in teams [e.g., 3]. Pirolli and Card's notional model of sensemaking [2] was initially applied to develop new technologies for intelligence analysts. Here the focus is specifically on tools that support the intermediate stages of extracting information, schematizing, and summarizing within a task force.

3 Requirements

In traditional collaboration settings, a benefit of forming a co-located task force was that the members would learn incrementally about each other and share content by working in close coordination, via face-to-face meetings and intermittent periods of asynchronous collaboration. A free benefit of working together was the effortless increment of mutual awareness and the common ground established, which would make members' coordination and sensemaking more efficient.

With the introduction of groupware applications, first, and web-based collaborative tools, later, collaborations in enterprises have become increasingly distributed in space and asynchronous in time. However, this greater flexibility in the setting and the ability to easily share large amount of data came with a big cost. In distributed, asynchronous collaboration, maintaining awareness and making sense of massive amount of content now requires both an active effort from the collaborators and adequate tools need to be provided within the shared workspace.

Nowadays, many tools (e.g., FriendFeed [4]) support task force members to collect low-level tokens of information such as web pages and Word documents into a shoebox-like repository (see left box in Figure 1). Similarly, several collaborative editing tools support the members at the end of the collaboration, while the report is being finalized (e.g., collaborative editing tools such as Google Docs or some features of Microsoft Office, see right box in Figure 1). But very little support is available for supporting awareness and sensemaking while the task force is engaged in the process

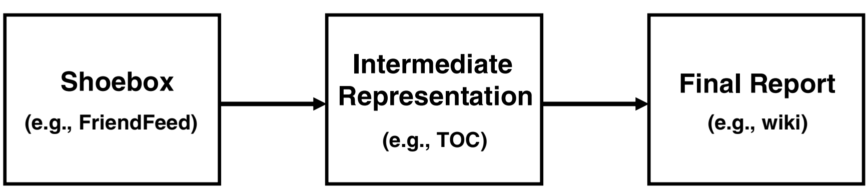


Fig. 1. Sensemaking process and sample tools: adaptation of model in [2]

of filtering out noisy collections of data to arrive at relevant information tokens, creating intermediate representations from the low-level tokens, and communicating the individual contributions to other people.

Proposition 1. Knowledge sharing. *In a task force, where collaborators have different roles, it is not the case that everyone in the team should know everything.*

This proposition emphasizes the need to reduce the things that collaborators need to attend to. This contrasts with the simplistic view of knowledge sharing as the formation of uniform mental models shared across team members (i.e. see this distinction in the literature on the transactive memory models [5]). It is, in fact, endemic to task forces dealing with massive amount of information to manage different skill sets and jobs. It is more efficient if the members divide the labor and attend to only what is relevant to their jobs. In this context, the performance can be improved via tools that support the awareness of each member for relevant content in ways that take into account the role specialization within the task force [e.g., 6].

Proposition 2. Content representation. *Collaborators assimilate shared content at a lower cognitive cost if they organize the large amount of information into higher-order content abstractions rather than low-order tokens or chunks of information (e.g., a few paragraphs of text, a spreadsheet or a table with data).*

This proposition points to a first solution to help with the management of very large amount of shared information (see proposition 1). It is quite common in large hierarchical organizations to generate briefings for the leaders that summarize large amount of data and detailed analyses (e.g., briefings for the US president).

Search engines, feed readers, and plain wikis allow the members of a task force to forage large amounts of detailed and unstructured data but provide little or no help for filtering, categorizing, and organizing the content. On the other hand, research on information processing and information visualization suggests that if a workspace integrates these foraging tools and in addition supports the construction of intermediate content abstractions, then it can significantly improve the quality of knowledge sharing and sensemaking [e.g., 7, 2].

Proposition 3. Role-specificity of representations. *The awareness and coordination of the collaborators improve if the abstractions also reveal about the roles of the authors.*

This proposition follows from the combination of propositions 1 and 2: if the co-workers are more efficient when they selectively attend to only what is relevant to them and if they can easily construct and share high-order representations of information, then it would be also helpful if such high-order representations can be personalized based on the members' roles (i.e., focusing on details relevant to each role) and carry also information about what role has contributed what content (i.e., supporting awareness of roles). Examples of features supporting awareness of different roles are provided in [3, 6].

Proposition 4. Content-argument proximity. *Sensemaking quality and motivation to contribute increase if the workspace presents the content (e.g., raw evidence) close to the rationale for sharing it (e.g., added arguments).*

This proposition points to the fact that the overall process of sharing, summarizing, and judging ideas in work groups has the general form of a dialog (e.g., initial proposal, reply, reply-to-reply, ..., deliberation). Moreover, when sharing a token of information that was already processed individually, collaborators tend to naturally attach their rationale or argument for sharing it [e.g., 8]. However, groupware applications (e.g., Groove) and wikis (e.g., MediaWiki applications) are designed with an unnecessarily marked separation between the content shared (documents or pages) and the discussion on it (i.e., discussion tools in groupware systems or discussion pages in wikis). This imposes extra steps (i.e., clicks) and context switches when users need to match the content with its rationale (and who contributed what).

In contrast, other web tools (i.e., blogs, forums) or groupware prototypes designed to make arguments both visible and visually related to the shared tokens have been successful in enabling high-quality sharing [9] and understanding [12], and high participation (see FriendFeed [4] or web blogs such as TechCrunch and Slashdot).

4 Workspace Design

In relation to the scenario above, this section illustrates the design of a workspace that helps to channel information from the foraging tools to the editing tool (i.e., the wiki) used by the task force. Each member has a personal and a shared space where the numerous pieces of information found can be pre-processed individually and then analyzed collaboratively (see Notebook in SparTag.us tool [15]). The filtered and commented content is then summarized and assembled in the wiki.

To address the requirements synthesized in the propositions above, the design includes support for content abstraction (proposition 2), selective awareness of shared content and contributors (propositions 1, 2, 3), discussion in context (propositions 2, 4) and guided discovery (propositions 1, 2, 3). Providing such support requires an adaptive workspace that models the knowledge, role, and interaction of each member and then provides support that is informed by each user model.

4.1 Modeling the Task Force at Work

Past work on adaptive systems (e.g., handheld guides in museums or online recommender systems) contributed sophisticated approaches to model individual users and guide their exploration of large amount of content [13, 14]. Such models allow tailoring the presentation on user's knowledge, interest, and interaction history and enabling personalized recommendations that guide the discovery of new content.

We adapt the model for individual users proposed in [13] and extend it to the case of a task force (Figure 2). The proposed model has four modules that keep track of four different sets of attributes for each member:

- Static member characteristics: e.g., age, gender, interface preferences
- Three sets of dynamic characteristics:
  1. *Individual knowledge* (i.e., facts s/he knows in the problem domain)
  2. *Role* (i.e., responsibilities in the task force and personal interests)
  3. *Interaction history* (i.e., content searched, content found, annotations, and summaries generated).

All the four modules of the model are initialized with information from users profiles (member-specified or imported from pre-existing task forces) and the roles assigned by the team leader of this task force. Then, the three dynamic modules are incrementally refined as the members work on the task.

A first novel element of this model is the account for the members' roles, which makes it a model for a work group. When considered collectively the different roles represent the strategy of the group. Key interdependencies among the roles can be inferred and used. This relates to recent attempts in collaborative computing to model the structure of collaborative activities or business projects rather than just their actors (activity-centric design [17]). A second novel aspect is that the model tracks no only the behavior of retrieving existing information but also the results of generating syntheses (written summaries), which includes content added ex-novo by members. Finally, in contrast with black-box modeling approaches, we propose a "see-through model": the facts collected by the system on the members are made visible to the members. This aim at enhancing their mutual awareness at a project level.

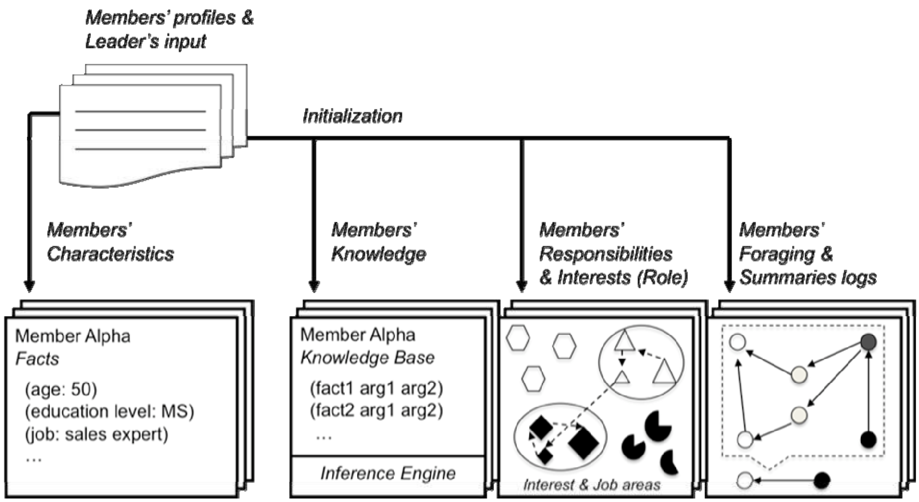


Fig. 2. Task Force Model Components

4.2 User Interface Functions

Let us consider the situation of a scientific task force sharing, discussing, and organizing information for a report on climate change. The leader has given the members different roles based on their expertise. The members pooled in a shared web-based repository a large number of pieces of content from papers, books, and web pages. These are presented in a list, which can be rearranged or filtered. Each piece can be annotated (e.g., highlighted, tagged).

To illustrate the functions of the user interface we use the table of contents (TOC) as our example content abstraction to organize and share information. A TOC is an intermediate representation on the content that is typically exploited in collaborative report-writing tasks. Note however that it is not the only possible embodiment of content abstraction. Alternative kinds of representations are clouds of tags, concept maps, affinity diagrams, flow charts, box-and-arrow (or Entity Relationship) diagrams, etc. The aim of this paper is to present the design ideas underlying the modeling approach and the UI functions, rather than the specific TOC representation.

We propose an augmented version of a TOC where the members can refine the items, access the low-level tokens in the leaf nodes, browse new low-level tokens selected by the system, and look at the contributions of other members, etc.

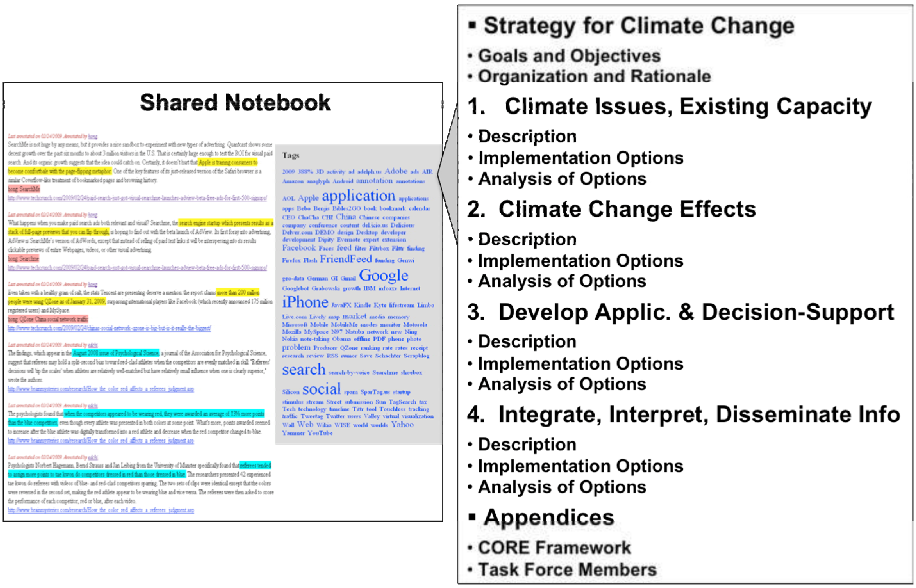


Fig. 3a (left) and 3b (right). 3a shows the shared repository with annotated pieces of content and tag cloud (functions implemented in SparTag.us [15]). 3b shows the proposed intermediate shared representation: a table of contents of the task force report.

4.2.1 Creating Intermediate Representations (Content Abstraction Function)

Group members select and categorize the low-order chunks of content from documents in the repository. Then they create items in the TOC (see proposition 2). Purpose-tagging can be used by the collaborators either while foraging [10] or later when sharing to explain why added tokens are relevant, or why contributed summaries are useful [8]. This enables clustering of content by purpose or argument (see proposition 4).

4.2.2 Selective Awareness of the Knowledge That Was Shared (Awareness Function)

More importantly, the TOC is displayed to each member as a personalized view (point-of-view) that takes into accounts the member's prior knowledge, role, and prior activity in the system (see proposition 3). Visual cues used to personalize the view are: expanding vs. collapsing parts of the table, boldening vs. graying out items, adapting font size of items, and showing visual traces of amount of overall content

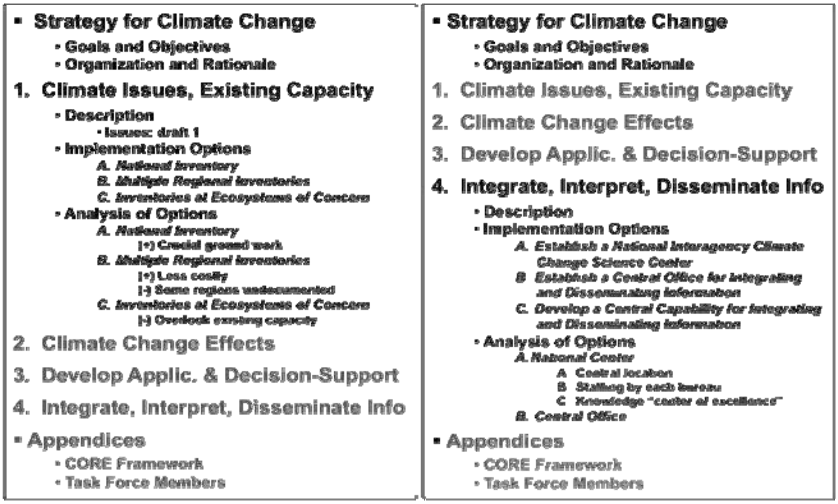


Fig. 4a (left) and 4b (right). Role-specific views of TOC shown in Figure 3a. For two experts who are focusing on, respectively, sections 1 and 4 of the report.

(overview on content) and other members' activity on the different parts of the table (activity progress bar). Figure 4a and 4b present views of the TOC to two members: Climate Issues and Capacity expert and Information Dissemination expert.

4.2.3 Supporting Discussion in Context (Discussion Function) and Scaffolding the Foraging of New Knowledge via Recommendations (Discovery Function)

A dialog or discussion concerning an item of the TOC can be conducted in context as shown in Figure 5a (see proposition 4). The contributor of an item can explain why he chooses to add the item. In addition, as shown in figure 5b, the system leverages the model (i.e., the information about prior activity of each member) to filter recommendations of new content or notification of related contributions. The member can adjust

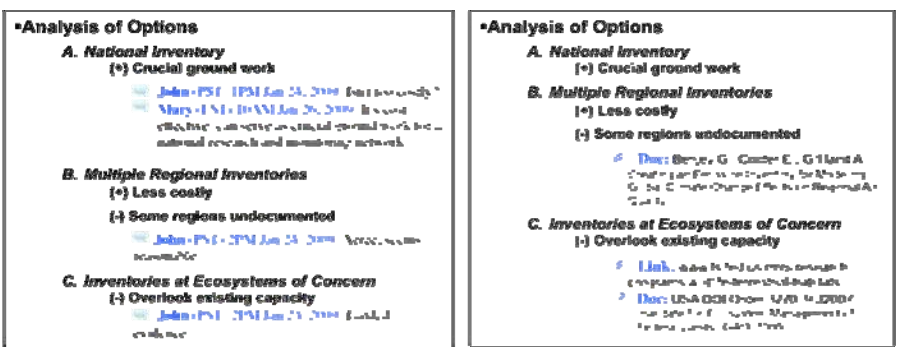


Fig. 5a (left). and 5b (right). 5a shows discussion messages in context. 5b shows three items recommended by the system. The icons indicate that two were rated as relevant.

the weight that the model has accumulated to influence the recommendations and notifications.

5 Discussion and Future Work

The sections above motivated and presented the design of a shared workspace. First, the approach to model knowledge, roles, and past contributions of the members, which are incrementally defined as they continue working together. Then, for the user interface (UI), the design of the workspace includes functions for (1) constructing intermediate representations to abstract and share knowledge efficiently, (2) selective awareness of what relevant knowledge was shared and who contributed it (3) discussion in context on the representation and discovery of new knowledge guided by notification of related contributions. These functions require a workspace that adapts to the needs of each member. This requisite motivates the modeling functions.

We presented design mockups and ideas to illustrate these UI functions. The proposed design builds on prior studies conducted in the Augmented Social Computing Area at the Palo Alto Research Center. Prior work has provided us with a web tool supporting individuals and groups at an early stage of sensemaking while they collect information, share, and learn from collaborators (see SparTag.us prototype [15] and study [16]). A precursor of the adaptive representations (TOCs) in this paper is the ScentIndex UI technique [18], which supports individual information foraging from a book via an enhanced subject index that reorganizes the content to suit the user's information needs. In other research we studied computer-supported teams at their final stage of sensemaking, while a final complex decision was made (see CACHE prototype and study [3]). Currently, we are extracting requirements by observing the work of real task forces. We have been observing expert professionals in an enterprise who take part in task forces, such as corporate teams that write business proposals on behalf of their company for competing in large bids.

As part of our future research, we plan to iteratively develop the design and perform formative evaluations with members of the task forces that are currently being observed for requirement elicitation. Consistently with our two design goals the evaluation measures include measures of process (costs for content abstraction and costs of awareness for content & roles) and measure of performance (quality of the report and total coverage of relevant information).

References

1. `http://www.usgs.gov/global_change/doi_taskforce.asp` (accessed on 02/25/2009)
2. Pirolli, P., Card, S.: The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis. In: Proceedings of 2005 International Conference on Intelligence Analysis, pp. 2–4 (1995)
3. Convertino, G., Billman, D., Shrager, J., Masur, J.P., Pirolli, P.: The CACHE Study: Group Effects in Computer-Supported Collaborative Analysis. Computer Supported Cooperative Work 17, 353–393 (2008)
4. `http://friendfeed.com/` (FriendFeed tool, accessed on 02/25/2009)

5. Mohammed, S., Dumville, B.C.: Team mental models in a team knowledge frame-work: expanding theory and measurement across disciplinary boundaries. Journal of Organizational Behavior 22 (2001)
6. Convertino, G., Mentis, H., Rosson, M.B., Slavkovic, A., Carroll, J.M.: Supporting Content and Process Common Ground in Computer-Supported Teamwork. In: Proceedings of CHI 2009 (to appear, 2009)
7. Resnikoff, H.L.: The illusion of reality. Springer, New York (1989)
8. Marshall, C.: Annotation: From paper books to the digital library. In: Proceedings of the ACM Digital Libraries 1997 Conference, Philadelphia, PA, July 1997, pp. 131–140 (1997)
9. McCarthy, J.C., Miles, V.C., Monk, A.: An experimental study of common ground in text-based communication. In: Proceedings of CHI 1991, pp. 209–215. ACM, New York (1991)
10. Berlin, L.M., Jeffries, R., O'Day, V.L., Paepcke, A., Wharton, C.: Where Did You Put It? Issues in the Design and Use of A Group Memory. In: Proc. INTERCHI 1993, pp. 24–29 (1993)
11. McDonald, D., Ackerman, M.: Collaborative Refinery: A Collaborative Information Workspace for the World Wide Web. Technical report 97-03, Dept of Information and Computer Science, University of California at Irvine (1997)
12. Singh, V., Twidale, M.B.: The Confusion of Crowds: non-dyadic help interactions. In: Proceedings CSCW 2008, pp. 699–702 (2008)
13. Sarini, M., Strapparava, C.: Building a user model for a museum exploration and information-providing adaptive system. In: Brusilovsky, P., De Bra, P. (eds.) The 2nd Workshop on Adaptive Hypertext and Hypermedia, Report No. 98/12, CS Reports, Eindho-ven University of Technology, Eindhoven, The Netherlands, pp. 63–68 (1998)
14. Pu, P., Viappiani, P., Faltings, B.: Increasing user decision accuracy using suggestions. In: Proceedings CHI 2006, Montréal, Québec, Canada, April 22-27 (2006)
15. Hong, L., Chi, E.H., Budiu, R., Pirolli, P., Nelson, L.: SparTag.us: A low cost tagging system for foraging of web content. In: Proc. AVI 2008, pp. 65–72. ACM, New York (2008)
16. Nelson, L., Pirolli, P.L., Hong, L., Chi, E.H., Held, C.: Examining the impact of social annotations in sensemaking tasks. Note submission in review for CHI 2009
17. Moran, T.P., Cozzi, A., Farrell, S.P.: Unified Activity Management: Supporting People in e-Business. Communications of the ACM 48(12), 67–70 (2005)
18. Chi, E., Hong, L., Heiser, J., Card, S.: ScentIndex: Conceptually Reorganizing Subject Indexes for Reading. In: Proc. IEEE Symposium on Visual Analytics Science and Technology (VAST), October 2006, pp. 159–166 (2006)

Use of Deception to Improve Client Honeypot Detection of Drive-by-Download Attacks

Barbara Endicott-Popovsky[1], Julia Narvaez[1], Christian Seifert[2], Deborah A. Frincke[3], Lori Ross O'Neil[3], and Chiraag Aval[1]

[1] University of Washington
4311 – 11th Avenue NE, Suite 400, Seattle, Washington 98105, USA
{endicott,jnarvaez,chiraaga}@u.washington.edu
[2] Victoria University of Wellington
School of Engineering and Computer Science, Victoria University
PO Box 600, Wellington 6140, New Zealand
christian.seifert@gmail.com
[3] Pacific Northwest National Laboratory
902 Battelle Boulevard, Richland, WA, USA
{deborah.frincke,lro}@pnl.gov

Abstract. This paper presents the application of deception theory to improve the success of client honeypots at detecting malicious web page attacks from infected servers programmed by online criminals to launch drive-by-download attacks. The design of honeypots faces three main challenges: deception, how to design honeypots that seem real systems; counter-deception, techniques used to identify honeypots and hence defeating their deceiving nature; and counter counter-deception, how to design honeypots that deceive attackers. The authors propose the application of a deception model known as the deception planning loop to identify the current status on honeypot research, development and deployment. The analysis leads to a proposal to formulate a landscape of the honeypot research and planning of steps ahead.

Keywords: deception, counter-deception, honeypots, drive-by-downloads, cyber-attacks.

1 Introduction

With increasing reliance on computer networks, important expected security concepts—confidentiality, integrity and availability—are under constant threat: 1) personal information, such as names/credit card numbers, is stolen; 2) office desktop computers are compromised into sending e-mail spam; and 3) risk of power grid outages caused by denial-of-service attacks on SCADA systems [1] might escalate.

A particularly insidious type of online attack has emerged in recent years, which targets clients through malicious servers that deliver an attack as part of the server's response to a client request. As the web browser requests content from a web server, the server returns a malicious page that launches a so-called drive-by-download attack on the browser. If successful, the web server pushes and executes arbitrary programs on the client machine.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 138–147, 2009.

Security devices called high-interaction client honeypots are able to find these malicious web pages by driving a client to visit web pages and make an assessment as to whether the page launches an attack. However, if the malicious server can first identify the client as a honeypot, it could choose not to launch attack code, rendering the client honeypot ineffective. Attacker counteracts are exemplified by articles on honeypot detection, in which several ways to fingerprint honeypots are introduced [2].

These researchers have concluded that the use of detection techniques in drive-by attacks necessitates the inclusion of deception techniques in client honeypots. With an understanding of the anti-detection techniques used by malicious servers, this paper proposes deception methodologies designed to develop client honeypots that elude detection. As the adversary improves in sophistication, so do the defenders.

2 Background

"A honeypot is a security resource whose value lies in being probed, attacked, or compromised". Even though the notions of honeypots were originated in the early 1990's, only recently commercial products have been developed and papers have been published [3]. The concepts of honeypots were formulated in 1990/1991 with the work of Clifford Stoll's ”The Cuckoo's Egg" and Bill Cheswick's "An Evening With Berferd" [4]. The use of honeypots and decoys as a deception in the defense of information systems was related by Cheswick, Bellovin, D'Angelo and Glick, in 1991 [5] in the paper "An Evening with Berferd In Which a Cracker is Lured, Endured, and Studied." The paper is a chronicle of how the researchers offered a “bite” to a cracker, the traps used to lure and detect him, and the chroot “Jail” the researchers built to watch his activities [6].

Types of honeypots can be differentiated by their ability to interact with an attacker. Systems that emulate vulnerabilities and allow limited interaction with the attacker are low-interaction honeypots. Systems that are vulnerable and allow interaction with the attacker at all levels are high-interaction honeypots [7]. Another differentiation is between physical honeypots, which run on physical machines, and virtual honeypots, which run on virtual machines [2].

As a result of attackers exploiting vulnerabilities in client programs (such as browsers), honeypots have evolved to simulate the behavior of a human and analyze how such behavior is exploited by an attacker [2].

2.1 Client Honeypots

A client honeypot consists of three components: the queuer, visitor, and analysis engine (Fig. 1 illustrates components). This client is controlled by a visitor component which interacts with potentially malicious web servers. Information about what server to interact with and the data to be sent to the server is created by a queuer component, for example a web crawler, that generates server requests. Lastly, the analysis engine assesses whether the server is malicious or benign.

The visitor component maps to high- and low-interaction client honeypots. The former allows the honeypot system full functional interaction. As the client interacts with the server, the system monitors for unauthorized state changes, such as file

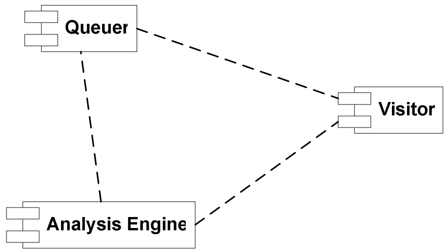


Fig. 1. Client Honeypot Component Diagram

modifications or process adjustments that would indicate a successful attack [8,9]. The latter signifies that the functionality of the client is limited, typically by using emulated services. Because no active exploitation occurs, the low-interaction client honeypot inspects the response directly using signatures, heuristics, and security predicates to detect attacks [9, 10, 11].

Given that honeypots are deceptive by nature, there is a wealth of wisdom to be gained from the study of deception theory in other sciences, such as social science.

2.2 Deception

The Longman Dictionary of American English defines Deception as "An act of deceiving." Deceiving is defined as "To cause someone to accept as true or good what is false or bad [13]." Multiple studies and theories of deception have been proposed. Cohen states that "Deception exploits errors in cognitive systems for advantage. It is achieved by systematically inducing and suppressing signals entering the target cognitive system [5]."

Bell and Whaley studied the general theory of deception and types of deception [14]. They argue that there are two levels of basic deceptive methods found in nature: hiding and showing. Humans consciously use these two methods found in nature.

Hiding, level one, is divided into three parts: masking, repacking and dazzling. Masking: the real thing is hidden by blending with the background, integrating itself with the surroundings, or seeking invisibility. Repacking: the real thing is perceived in various ways, as dangerous, harmless or irrelevant. Dazzling: ultimate problem of what to do when masking and repacking do not work and the attacker knows the victim is there. The qualities of the object might be changed as to confound [14].

Showing, level two, is divided into three parts: mimicking, inventing and decoying. Mimicking: a replica of reality is created by selecting one or more characteristics of the real in order to achieve an advantageous effect. Inventing: the false is presented through the creation of an alternative reality, e.g. the false document appears to be real, but it is not. Decoying: gives an additional alternative pattern, increasing its certainty [14]. The work performed by honeypots fits within these levels and categories of deception.

3 Problem: Detecting Honeypots

The design of honeypots faces three main challenges: deception, counter-deception and counter-counter-deception [15]. a) Deception problem: how to design honeypots that look like normal computer systems. b) Counter-deception problem: techniques used to identify if a computer is a honeypot. Objectives of counter-deception include the appraisal of whether an attacker can detect a honeypot, and the identification of whether the data collected from such a honeypot are misinformation. c) Counter-counter-deception: how to design honeypots that make attackers think that they are real systems [15].

4 Analysis

Honeypots are used to research and to prevent, detect, and respond to attacks. For research purposes, honeypots collect information on threats, which can be used for trend analysis, identification of new tools or methods, and attacker identification [16]. In this section, the authors focus the analysis on the research purpose of honeypots.

4.1 Deception

Before launching an attack, adversaries collect information about the host operating system and services running. Learning about the operating system allows attackers to understand what vulnerabilities the host might have. Learning about the services and versions facilitates planning of a route of attack [2]. Researchers value the knowledge of how the adversary breaks into a target machine and honeypots enable them to do that. The type of honeypot used varies according to the intended victim of attacks, which can be targeted attacks or targets of opportunity.

Targeted attacks are directed to targets of choice, which are organizations with high value information resources. For these targets of choice, production honeypot file servers could be used to provide falsified information to a human attacker who analyzes information given out by the honeypot [4]. Creating fake file systems is a form of mimicking and inventing [15]. Spitzner proposes the use of honeytokens, which are digital information entities, not computers. Any interaction with them is an unauthorized interaction. This form of honeypot is also useful to detect, identify and gather information about the malicious insider threat [17].

Targets of opportunity attacks can use multiple deception techniques, e.g. honeypot farms, in which honeypots are services. All the traffic coming to the production server is re-routed to pass through honeypots that are locally or remotely located. The honeypots need to emulate the production systems. In the event of detecting malicious activity, this can be logged, trapped, and traced back [18]. Roaming honeypots are mechanisms that allow the locations of honeypots to be unpredictable, continuously changing, and disguised within a server pool, from which a subset of servers provides services and the rest of the server pool is idle and acts as honeypots [19].

Client honeypots simulate the behavior of a human and actively search for attacks and malicious content on the Internet [2]. The level of interaction between client honeypots with servers can be low or high. Low-interaction client honeypots use a simulated client in place of a browser and assess the malicious nature of a server via static analysis such as signatures. High-interaction client honeypots interact with servers and assess the malicious nature of the server based on state changes [7].

Significant development of client honeypots is expected for web clients, the most critical of the cross-platform vulnerabilities in the SANS Top 20 list. Honeypots for newer applications such as VoIP and SCADA may become widespread [7].

4.2 Counter-Deception

Malware is increasingly more sophisticated. Developers of malware aim to make it undetectable. New offensive techniques are adopted once they are made public and quickly adapted to face new defensive techniques [21].

Examples of counter-deception are found in publications in Phrack magazine describing methods to detect, disable, and defeat Sebek [1] [22] in an attempt to avoid malware collection and hence malware analysis.

A trend has emerged in which malware uses evasion, e.g. the Agobot botnet family uses polymorphism as an obfuscation mechanism [20]. Malware is able to detect whether it is running in a virtual machine and change its behavior, e.g. a specimen discovered by Intelguardians [12], the worm Conficker, the Storm worm [24], and Agobot [23]. Examples developed by security researchers include Nopill [26], Vmdetect [27], Redpill [28], Scoopy Doo [2], and VMwareTools [29]. Scientific literature on the topic of detecting honeynets includes *NoSEBrEaK - Attacking Honeynets*, by Dornseif et al. who demonstrate methods to control honeynets [22].

In [30] two broad groups of strategies for detecting deception were identified: strategies based on detection of evidence of deception in the environment, and inspection for signs of deception in the information within the environment.

Honeypot detection methods usually exploit discrepancies between the real systems and honeypots [2]. Provos and Holz discuss several techniques to detect low-interaction and high-interaction honeypots. Realistic looking low-interaction honeypots need to deceive network scanning tools. High-interaction honeypots need to simulate an entire operating system environment. The deceiving nature of physical high-interaction honeypots can be concealed; however, honeypots running in virtual environments have additional challenges as virtualization is detectable [2].

Methods of virtualization detection exploit logical discrepancies, resource discrepancies and timing discrepancies. a) Logical discrepancies evaluate semantic differences in the interfaces of real and virtual hardware. b) Resource discrepancies evaluate the resources that the virtual machine shares with its guests, such as CPU cycles, physical memory and cache footprint. c) Timing discrepancies evaluate the variance in latency, relative differences in the latency of any two operations, and the behavior of these latencies over time [31]. The main reason for these discrepancies is that the virtual machines were designed to provide fidelity, performance and safety, but not transparency [2].

Several methods suggest themselves for detecting client honeypots. a) Observing click rate and dwell time could identify a client honeypot tasked with identifying malicious web pages as fast as possible. b) Referrer evaluation is another mechanism that identifies client honeypots based on their navigational characteristics. c) Another possible identification means is the network location of the incoming requests. These techniques are applicable to both low- and high-interaction client honeypots.

---

[1] Developed by the Honeynet Project [25], Sebek is a tool for collecting forensic data from compromised high-interaction honeypots [2].

There are other techniques that are specific to this type of client honeypot. High-interaction client honeypots could be identified by rendering checks. As a page is loaded, an adversary could check whether the page is actually displayed.

A low-interaction client honeypot likely appears like a regular browser. Using the header fields of entire requests can uncover this deception. Header order and data formatting might also give the deception away or the TCP/IP track can be analyzed with the passive OS fingerprinting tool p0f [32].The header lines and values of an http request can be analyzed and compared to a fingerprint database to identify a given web browser using browserrecon [33]. Further, low-interaction client honeypots are light weight, stripped-down versions of the browser. An adversary can discern it by calling functionality that is present in a full-fledged browser, but not in a low-interaction client honeypot [34]. Because low-interaction honeypots simulate a system and do not provide a complete operating system environment to the adversary, they can be detected more easily than high-interaction client honeypots.

4.3 Counter Counter-Deception

These researchers refer to counter counter-deception as the analysis of attackers' counter-deception techniques that result in new deception designs. The changes occurring in Sebek's code after publication of *Advanced Honey Pot Identification* [22], describing methods to defeat Sebek, is an example of counter counter-deception.

Counter counter-deception focuses on two main areas: creation of defenses and understanding how attackers work and think. The authors believe that this understanding will lead to improvements in honeypot research and development, applying deception techniques. Seifert, et. al., proposed a taxonomy of honeypot systems that facilitates the understanding of honeypot technology by presenting a faceted classification that addresses six areas of honeypot study: interaction level, data capture, containment, distribution appearance, communication interface and role in multi-tier architecture. This taxonomy offers a framework for describing honeypot research [35]. The values for each area [35] are shown in Table 1.

Table 1. Honeypot Taxonomy

Category	Interaction Level	Data Capture	Containment	Distribution Appearance	Communication Interface	Role in Multi Tier Architecture
Values	-High -Low	-Intrusions -Events -Attacks -None	-Defuse -Block -Slow Down -None	-Distributed -Stand-Alone	-Software API -Network IP -Non Network Hardware IF	-Client -Server

The authors argue that the systematic application of Bell and Whaley's theory of deception, using the taxonomy of honeypots, facilitates the identification of potential research gaps. According to Bell and Whaley, even though most cheating is done intuitively, the complex process to plan and design a deception can be depicted in a Deception Planning Loop. Deception falls in categories within two levels, hiding and showing [14], as shown in Table 2.

Table 2. Deception Levels and Categories

Level	**Hiding**	**Showing**
Category	- Masking - Repacking - Dazzling	- Mimicking - Inventing - Decoying

These categories give a spectrum of characteristics or charcs [14] (e.g. taxonomy of honeypots) to be used during the deception. The ruse is the process of selecting the appropriate categories of cheating and subsequently the characteristics to create a cover or effect. Ruses fall in categories: unnoticed, benign, desirable, unappealing and dangerous. The ruse creates a cover or effect for the attacker to accept the illusion. The planning of the deception aims at anticipating the illusion; however, the illusion depends only on the perception of the target audience [14].

Bell and Whaley describe the *Deception Planning Loop* as:

> *…Fashioning a RUSE from CHARCS that are projected by a selected CHANNEL as an EFFECT or COVER that, if successful, created an ILLUSION made up of the perceived CHARCS that is, therefore, a successful stratagem supporting the Deception Goal and hence the Strategic Goal* [14].

The deception model varies for attacks focused on targets of choice, and attacks focused on targets of opportunity executed with automated tools such as worms [4]. It also varies according to the type of honeypot. High-interaction honeypots are examples of mimicking users, browsers, and active content. Low-interaction honeypots are examples of decoying.

For example, the deception goal of a high-interaction client honeypot is to "look" like a human user and be attacked. The ruse is to mimic the human behavior by navigating on the Internet (charc/channel) and interacting with servers using a web browser (charc/channel). The operating system and applications have a degree of known vulnerabilities that are controlled according to the empirical experiment (charc/channel). If successful, the malicious server will have the illusion that the client honeypot is an actual user and will execute the attack.

The definitions of new approaches to develop honeypots are examples of different ruses. Some new approaches to develop honeypots have been formulated. For example, Vukasin Pejovic et al. conducted an initial investigation and implementation steps for the deployment of honeypots as an independent hardware device with the incorporated honeypot behavior [36].

These researchers argue that for the future development of honeypots, the results of a deception plan should retro-feed the *Deception Planning Loop*, making the definition of charcs and channels an ongoing process. For example, attackers frequently use compromised computers to spread attacks. To prevent these attacks, using deception techniques, honeypots control the data leaving them. E.g. Sebek and other Gen II honeynets impose a hidden limit to the number of outbound connections [37]. Lessons learned from some experiments are useful when planning the deception. For instance, Rowe and Goh observed increasing number of attacks after the system went down and came back up. This analysis suggests that keeping an existing long-used IP address and responding normally to packets might lead to a decrease in the number of attacks [38].

The authors believe that counter counter-deception in the development of client honeypots, in addition to a technical approach, should be complemented by a political and social approach to learning about trends and alerts in the attacker community. This is part of a framework to study malware, attackers' behavior and attack trends.

The Honeynet Project deployed the Global Distributed Honeynet project, with goals such as global deployment of more high-interaction honeynets, and cross referencing of incident data for correlation against historical forensic databases [39].

The NOAH Project, funded by the European Commission, is a three year project that intends to gather and analyze information about the nature of Internet cyber attacks [40]. Its *Honey At Home* implementation project extends its network to homes and small businesses [41]. It will develop an infrastructure to detect and provide early warning of attacks to expedite countermeasures to combat them.

Fred Cohen proposes the creation of a set of red teaming experiments in which attackers as well as defenders are studied, to understand how attackers work and think, and the effects of defenses on attackers [42]. Moreover, to isolate the effects of deception, he proposes the creation of control groups, and experiments with double blind data collection [42].

5 Conclusions

The determination of the current status of honeypot research and deployment by using deception theory can help identify which areas of honeypot technology research are priorities. This would be part of a framework to analyze malware, attacks and attackers' trends.

Stating the strategic deception goals, studying the feasibility of application of deception techniques available in the social sciences, becoming aware of what technology is available and the research status of such technology, and assessing the level of accomplishment of goals, would guide the depiction of the honeypot research and deployment landscape in order to indicate future research direction.

These researchers believe that aggregation, sharing and analysis of data captured with honeypots help describe the status of attacks and attacker trends. Adopting a taxonomy of honeypots enables the research community to agree on the object of study and facilitates needed communication.

References

1. Finisterre, K.: The Five Ws of Citect ODBC Vulnerability CVE-2008-2639 (2008), `http://www.milw0rm.com/papers/221`
2. Provos, N., Holz, T.: Virtual Honeypots: From Botnet Tracking to Intrusion Detection. Pearson Education, Boston (2008)
3. Tan, P., Kumar, V.: Discovery of Web Robot Sessions Based on their Navigational Patterns. Data Mining and Knowledge Discovery 6(1), 9–35 (2002)
4. Spitzner, L.: Honeypots: Tracking Hackers. Addison-Wesley, Boston (2003)
5. Cohen, F.: The Use of Deception Techniques: Honeypots and Decoys. In: Bidgoli, H. (ed.) Handbook of Information Security, vol. 3, pp. 646–655. John Wiley & Sons, Chichester (2006)

6. Cheswick, B.: An Evening with Berferd in which a Cracker is Lured, Endured, and Studied (1991), http://www.cheswick.com/ches/cv/main.html
7. Riden J., Seifert C.: A Guide to Different Kinds of Honeypots. Security Focus (2008), http://www.securityfocus.com/infocus/1897/3
8. Seifert, C., Steenson, R.: Capture - Honeypot Client. Honeynet Project (2006), https://projects.honeynet.org/capture-hpc
9. Moshchuk, A., Bragin, T., Gribble, S.D., Levy, H.M.: A Crawler-based Study of Spyware on the Web. In: 13th Annual Network and Distributed System Security Symposium, The Internet Society, San Diego (2006)
10. Seifert, C., Welch, I., Komisarczuk, P.: HoneyC - The Low-Interaction Client Honeypot. In: NZCSRCS, Hamilton (2007), http://www.mcs.vuw.ac.nz/~cseifert/blog/images/seifert-honeyc.pdf
11. Ikinci, A., Holz, T., Freiling, F.: Monkey-Spider: Detecting Malicious Websites with Low-Interaction Honeyclients. In: Sicherheit, Saarbruecken (2008)
12. Carpenter, M., Liston, T., Skoudis, E.: Hiding Virtualization from Attackers and Malware. IEEE Security & Privacy 5(3), 62–65 (2007)
13. Longman: Dictionary of American English. Longman, White Plains (1983)
14. Bell, J.B., Whaley, B.: Cheating and Deception. Transaction Publishers, Edison (1991)
15. Rowe, N.C.: Measuring the Effectiveness of Honeypot Counter-Counterdeception. In: Proc. of the 39th Hawaii International Conference on System Sciences, vol. 6, p. 129c. IEEE Xplore (2006)
16. The Honeynet Project: Know Your Enemy: Learning About Security Threats. Pearson Education, Boston (2004)
17. Spitzner, L.: Honeypots: Catching the Insider Threat. In: Proc of the 19th Annual- Computer Security Applications Conference, pp. 170–179. IEEE Xplore (2003)
18. Lakhani, A.D.: A dissertation on Deception Techniques Using Honeypots. Information Security Group, Royal Holloway, University of London
19. Khattab, S.M., Sangpachatanaruk, C., Moss, D., Melhem, R., Znati, T.: Roaming Honeypots for Mitigating Service-Level Denial-of-Service Attacks. In: Proc. of the 24th ICDCS 2004, pp. 328–337. IEEE Computer Society, Washington (2004)
20. Barford, P., Yegneswaran, V.: An Inside Look at Botnets. In: Christodorescu, M., Jha, S., Maughan, D., Song, D., Wang, C. (eds.) Advances in Information Security, vol. 27, pp. 171–191. Springer, US (2007)
21. Harris, S., Harper, A., Eagle, C., Ness, J.: Gray Hat Hacking. McGraw-Hill, New York (2007)
22. Dornseif, M., Holz, T., Klein, C.N.: NoSEBrEaK - Attacking Honeynets. In: Proc. of the IEEE Workshop on Information Assurance and Security, pp. 123–129. IEEE Xplore (2004)
23. Dittrich, D.: VMWare Detection?. Virus.org (2004), http://lists.virus.org/honeypots-0411/msg00044.html
24. Zdrnja, B.: More Tricks from Conficker and VM Detection. SANS Internet Storm Center (2009), http://isc.sans.org/diary.html?storyid=5842
25. Spitzner, L.: Sebek. The Honeynet Project, http://www.honeynet.org/project/sebek
26. Quist, D., Smith, V.: Detecting the Presence of Virtual Machines Using the Local Data Table. Offensive Computing, http://www.offensivecomputing.net
27. Lallous: The Code Project. Detect if Your Program Is Running Inside a Virtual Machine (2005), http://www.codeproject.com/KB/system/VmDetect.aspx?display=Print
28. Rutkowska, J.: Red Pill or How to Detect VMM Using (Almost) One CPU (2004), http://invisiblethings.org/papers/redpill.html
29. Kato, K.: VM Back, http://chitchat.at.infoseek.co.jp/vmware/vmtools.html

30. Santos, E., Johnson, G.: Toward Detecting Deception in Intelligent Systems. In: Proc. SPIE the International Society for Optical Engineering, vol. 5423, pp. 130–141. SPIE, Bellingham (2004)
31. Garfinkel, T., Adams, K., Warfield, A., Franklin, J.: Compatibility Is Not Transparency: VMM Detection Myths and Realities. In: Proceedings of the 11th USENIX workshop on hot topics in operating systems (2007), `http://www.usenix.org/event/hotos07/tech/full_papers/garfinkel/garfinkel_html`
32. Zalewski, M.: The New p0f: 2.0.8 (2006-09-06), `http://lcamtuf.coredump.cx/p0f.shtml`
33. Ruef, M.: Browserrecon Project, `http://www.computec.ch/projekte/browserrecon`
34. Hoffman, B.: Circumventing Automated JavaScript Analysis. In Black Hat USA, Las Vegas (2008), `http://www.blackhat.com/presentations/bh-usa-08/Hoffman/Hoffman-BH2008-CircumventingJavaScript.ppt`
35. Seifert, C., Welch, I., Komisarczuk, P.: Taxonomy of Honeypots. Technical Report CS-TR-0. School of Mathematical and Computing Sciences. Victoria University of Wellington (2006)
36. Pejovic, V., Kovacevic, I., Bojanic, S., Leita, C., Popovic, J., Nieto-Taladriz, O.: Migrating a HoneyDepot to Hardware. In: The International Conference on Emerging Security Information, Systems, and Technologies, pp. 151–156. IEEE Xplore (2007)
37. Rowe, N.C.: Deception in Defense of Computer Systems from Cyber Attack. In: Janczewski, L.J., Colarik, A.M. (eds.) Cyber Warfare and Cyber Terrorism, pp. 97–104. IGI Global, Hershey (2008)
38. Rowe, N.C., Goh, H.C.: Thwarting Cyber-Attack Reconnaissance with Inconsistency and Deception. In: Proc. of the IEEE Workshop on Information Assurance United States Military Academy, West Point, NY, pp. 151–158. IEEE Xplore (2007)
39. Watson, D.: GDH Global Distributed Honeynet. The Honeynet Project (2007), `http://www.ukhoneynet.org/PacSec07_David_Watson_Global_Distributed_Honeynet.pdf`
40. European Network of Affined Honeypots: About NoAH, `http://www.fp6-noah.org/about`
41. European Network of Affined Honeypots: honey@home, `http://www.honeyathome.org`
42. Cohen, F.: The Use of Deception Techniques: Honeypots and Decoys, `http://all.net/journal/deception/Deception_Techniques_.pdf`

Capturing and Building Expertise in Virtual Worlds

Jared Freeman, Webb Stacy,
Jean MacMillan, and Georgiy Levchuk

Aptima, Inc., 1726 M Street, NW, Suite 900, Washington, DC 20036, USA
{jfreeman,wstacy,jmacmillan,glevchuk}@aptima.com

Abstract. Model-driven simulation can make the design and delivery of instruction more efficient and effective. We describe two computational models that support both the design and delivery of instruction. BEST (the Benchmarked Experiential System for Training) can guide experts through the space of domain problems during the knowledge engineering phase of instructional design; it can guide trainees through the space of training objectives during instruction. PRESTO (Pedagogically Relevant Engineering of Scenarios for Training Objectives) builds scenarios on the fly to elicit the knowledge of experts during instructional design, and to satisfy the instructional objectives of trainees.

Keywords: Adaptive instruction, knowledge engineering, Constraint Logic Programming, Markov Decision Process.

1 Introduction

Training organizations often marry 21st century simulation technology to 20th century methods of designing instructional content and training scenarios.

Instructional design, which defines the objectives and content of training, engage scarce experts in interviews and qualitative analyses that are so time-consuming they are derided as a bottleneck [1].

Scenario design in simulation-based training creates artworks – engaging but singular products that present fundamentally the same instructional experience to all students. Rarely are training scenarios developed in libraries that systematically vary factors related to training objectives, and rarely can simulators adapt scenarios on the fly to respond to the instructional needs of each student. This is true, in general, whether we conceive of scenarios as brief vignettes (e.g., a specific air-to-air engagement) or as the larger events in which they are combined (e.g., a mission of many engagements), and whether the training simulator serves a one trainee (e.g., a game) or many (e.g., a military exercise with live, virtual, and constructive components).

We are exploring ways in which simulations can be mated to computational models that specify and construct scenarios to meet their users' needs. This strategy has value during scenario design; it configures training scenarios to the needs of multiple students. The technologies also have value during instructional design, where they could improve the capture and analysis of expertise.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 148–154, 2009.

We open this paper with a brief discussion of expertise. It is what instructional designers seek and what scenario designers hope to deliver to students through their simulations. In particular, we develop a spatial metaphor for expertise, and we use this to describe how two modeling technologies that we have implemented automate scenario design and could, potentially, automate instructional design.

2 Expertise

Expertise is the ability to discriminate meaningful classes of domain features and patterns, and to take decisions or actions that are appropriate to the class at hand.

This conceptualization of expertise is illustrated in Figure 1, which depicts the novice's conception of a domain as an undifferentiated space of objects and events. In contrast, the expert's understanding is structured to distinguish functionally distinct and important objects, events, and situations. Expert partitioning of the domain space supports decision making and action. Each partition of the space is associated with distinct issues for consideration in decision making, or with unique actions such as tactics.

Studies of physicists, software engineers, chess masters, and others validate that domain knowledge is more functional (useful) among experts than novices, and more uniform. For example, expert physicists sort physics problems by the deep structure or functions they represent (e.g., acceleration), while novices sort by surface features such as the materials (e.g., balls rolling down hills) cited in the problem [2] Expert software engineers sort programming terms into groups that are similar to those of other software experts and are relevant to programming, while less expert engineers are more varied in their classification schemes and those schemes are often based on everyday meanings of the terms [3]. Chess masters reconstruct briefly viewed chess boards in tactically important clusters of pieces (e.g., defensive and offensive clusters), while novices rebuild boards in an order that is arbitrary, with respect the strategies of chess [4]. In short, expert knowledge of domains is partitioned in ways that reflect meaningful differences between problems, differences that bear on the accuracy of decisions and actions.

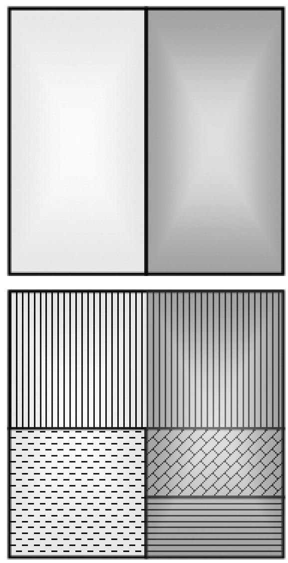


Fig. 1. Novices perceive and respond to surface features of domain problems (shading, above). Experts detect underlying problem structure (patterns, below), which maps well to correct solutions.

The concept that expert knowledge is functionally structured has two ramifications for the design of scenarios or their component vignettes. First, we should create the range of scenarios that represent each partition of the domain, so that students can experience the challenges they pose. Such scenarios must represent the training objectives students need to address, the conditions (e.g., events) under which students must learn to perform, and the measures with which trainers evaluate student expertise. Second, we must guide students from scenario to scenario (partition to partition) so that they learn to distinguish between the problems each presents, and respond to them as experts do.

This concept of expertise also has implications for using simulations as knowledge engineering tools during instructional design. First, we can build scenarios (or vignettes) that represent the full range of domain problems, to enable experts to demonstrate their capabilities. Second, we can guide experts through that space in a way that enables us to efficiently discover the bounds experts draw, the partitions. This capability is most useful in domains in which experts are known to exist, in which the structure of expert domain knowledge is unknown or contested, and actions or tactics are not well mapped to that structure.

We are developing two modeling technologies that address these opportunities to improve training. To deliver scenarios that meet the specific needs of each student, BEST (the Benchmarked Experiential System for Training) guides students through the space of instructional objectives and corresponding scenarios; this demonstrably accelerates the development of expertise. PRESTO (Pedagogically Relevant Engineering of Scenarios for Training Objectives) adapts scenarios on the fly to satisfy instructional objectives or specifications, such as those BEST recommends. These models also have the potential to break the bottleneck of knowledge engineering. BEST can pilot experts efficiently through a potentially vast space of domain problems, which PRESTO then composes in scenarios on demand. The clusters of problems to which experts respond similarly each define a partition, a subspace of functionally similar domain problems. Instructional designers, having discovered these partitions, can rapidly develop didactic training that emphasizes the characteristics that distinguish each partition (or cluster of problems). They can transform the scenarios used in knowledge engineering into training scenarios that give novices a truly representative sample of experiences in the domain. We describe each of these technologies, below.

3 Scenario Specification and Sequencing with BEST

The Benchmarked Experiential System for Training (BEST) specifies the training treatment that a simulation should present next to advance a student farthest towards expertise.

To accomplish this, BEST uses a Partially Observable Markov Decision Process (POMDP) model to represent our probabilistic knowledge of the level of student competency (in multiple dimensions), and the probabilistic effects of training treatments on these competencies. These training treatments may be expressed as training objectives ("The student should acquire skill X, next"), training conditions ("Increase the challenge from enemies."), or training scenarios ("Present scenario 127 next"). Their instructional effects of each treatment on students of different abilities can be estimated by experts or learned from training performance data. BEST computes an optimal training policy [5] that adapts over time to specify next most beneficial simulation scenario given the trainee's most recent as well as history of performance.

In two experiments to validate BEST as a training adaptation tool, teams of subjects received simulator training in the complex tasks of the Air Operations Center

Dynamic Targeting Cell. The training experiences were drawn from a large library of approximately 50 scenarios that varied systematically on the intensity of defensive and offensive challenges. The instructional strategy for selecting from this library was controlled either by the BEST POMDP or by a hierarchical part-task training scheme that advanced trainees through increasing numbers of targets (offensive challenges), and then increasing numbers of threats (defensive challenges). The BEST solution reliably increased learning relative to the control condition, holding the number of training trials constant (see Figure 2) [6].

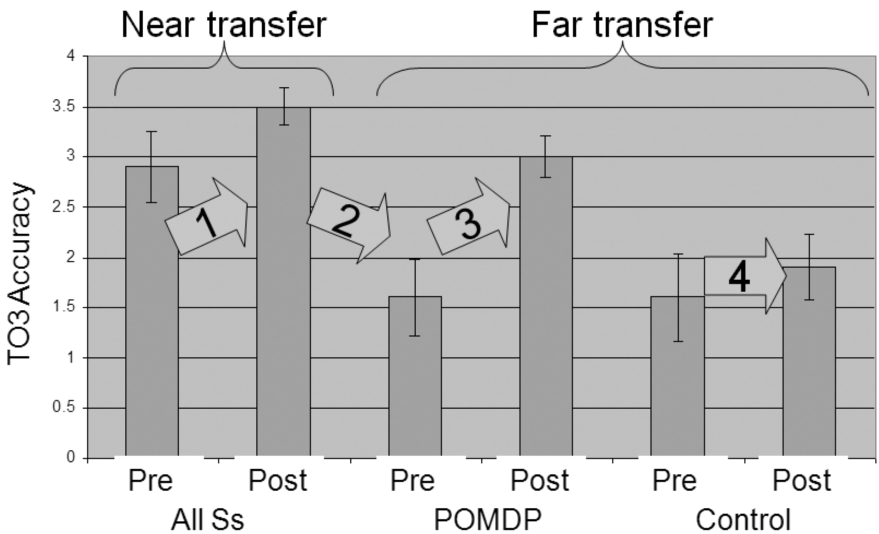


Fig. 2. Experimental tests of BEST found that teams learned the complex task of the Air Operations Center Dynamic Targeting Cell ($p<.01$); that far transfer within the task degrades performance ($p<.01$); that, on far transfer problems, teams in hierarchical part-task condition (control) do not reliably learn ($p>.05$), while those in the BEST / POMDP condition do ($p<.01$)

In the same way that BEST navigates a space of existing scenarios or training objectives, it can be used to learn the structure of an ill-defined domain space by observing expert behavior. This application of BEST, which we plan to attempt in the near future, requires that a large library of scenarios be constructed or generated on the fly (see PRESTO, below). These scenarios would sample the space of domain conditions randomly (at worst) or systematically according to some theory about important classes of domain problems or features. BEST would apply policy learning techniques to (1) formal descriptions of the features of each scenario and (2) expert decisions or actions in the scenarios. It would discover those partitions of the space (clusters of scenarios) to which experts generally respond in a uniform manner. Each such partition in the domain space consists, by definition, of functionally equivalent problems. In the terms of a Markov Decision Process, BEST would discover the groups of scenarios that constitute each unique "state", and it would discern the "action" that experts apply to the state to address the challenge it poses.

This application of BEST could be used to confirm theories about how experts understand and act in complex domains. In air-to-air warfare, the tactics of enemies are well known and our responses are well documented. However, experts make subtle distinctions in their interpretation of enemy approaches and their own application of tactics; model-driven simulations might efficiently discover these distinctions. BEST may have greater value in domains of human performance that are not well understood. It might, for example, be used to analyze how experts inspect multi-spectral imagery, how they monitor crowds to identify potential insurgents, and how they scan roadways for Improvised Explosive Devices. In all of these cases, experts perceive configurations of the environment as innocent, suspicious, or threatening; the take actions to test their perceptions; and they may take actions to address threats. To train the next generation of experts, we must analyze how today's experts understand (i.e., partition) the problem space in the domain, and how they choose their actions. Model-driven simulation should make this analysis much more thorough and efficient than traditional interview methods alone.

4 Scenario Construction with PRESTO

PRESTO (Pedagogically Relevant Engineering of Scenarios for Training Objectives) builds or revises simulation scenarios to ensure that they present the conditions required to meet specific training objectives [7, 8].

PRESTO accomplishes this by formally representing training objectives, domain objects, and domain events. PRESTO applies constraint logic programming (CLP) techniques to (re)formulate the schedule of training objectives and events addressed by a scenario. PRESTO defines temporal relations between domain events, spatial relations between domain objects that are required for training events, and other necessary scenario preconditions, so that it is possible for students to address training objectives (see Figure 3). It is particularly useful in simulations in which there are many students with potentially conflicting training objectives, all operating in a single, complex environment.

For example, suppose that a helicopter crew is training on a sensor fusion task. Performance measures indicate that they need additional work when certain combinations of targets are present. However, this combination occurs only once in the original scenario, and that occurrence was fouled when the crew unexpectedly chose a course that took them far away from the sensor targets, early the training scenario. In mid-exercise, PRESTO can determine whether it is feasible to add more instances of this target combination, and determine when and how to do so, given the constraints imposed by satisfying all other training objectives for all trainees engaged in the scenario. In reworking the scenario, PRESTO takes care not to interfere with the ongoing training of other students.

PRESTO is designed to support training delivery, per the description above. However, we hope to apply it to facilitate knowledge engineering as well. PRESTO can represent theories about the classes of problems experts perceive in a domain, and generate scenarios that present each class. When expert actions are relatively uniform across a group of scenario events (i.e., problems), a coherent partition of the space has

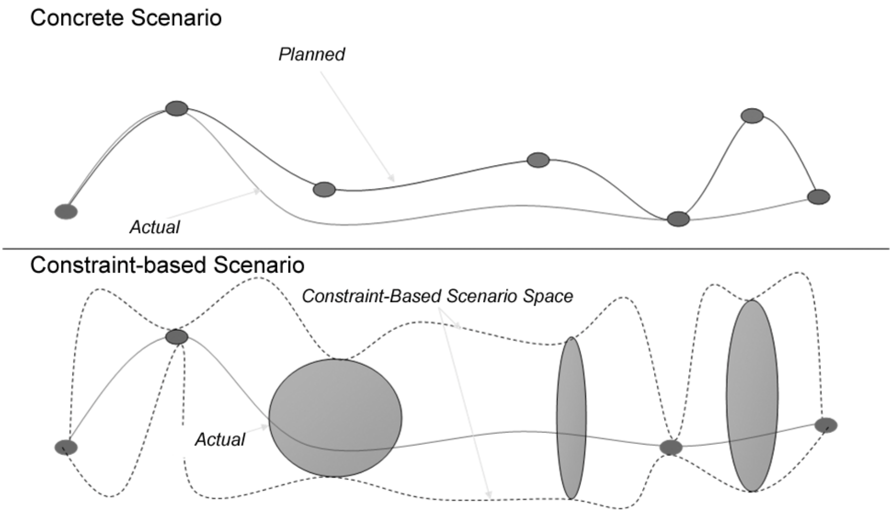


Fig. 3. Simulation scenarios are traditionally designed to draw trainees through a specific series of instructionally critical events (above), but these plans are often foiled by the actions of other trainees or system failures. PRESTO effectively models the bounds (constraints) of the simulation space and generates scenario events within those bounds to meet training objectives (below).

been discovered. When expert actions vary between scenarios, they represent meaningfully different problems or partitions of domain knowledge. PRESTO, in short, can generate representative samples of domain problems in a controlled manner, so that analysts can learn which problems are substantively different to experts.

5 Conclusion

We have described two computational models that can drive simulations to improve training delivery and knowledge elicitation. BEST specifies the training objective or class of problem from which a student can glean the greatest instructional benefit. PRESTO constructs scenarios that present such problems. These models were developed to ensure that simulators adapt to trainees, whether by selecting the best scenario among a library of many, or generating a scenario customized to the trainee. These technologies have potential, also, to accelerate knowledge engineering during instructional design. They should, if adapted to that task, help define and generate the content for training in new and complex domains. Model-driven simulation has the potential to make instructional design and scenario design more systematic and effective. This should greatly increase the impact of simulation-based training.

Acknowledgments. The research and development reported in this article were funded in part by the Department of Defense. We thank Melissa Walwanis of NAVAIR / Orlando for supporting development of PRESTO for automated scenario generation. We thank the Jerome Busemeyer and Robert Sorkin, formerly of the Air

Force Office of Scientific Research, for supporting the development of BEST algorithms and the empirical validation of them, which we conducted with Wayne Shebilske of Wright State University.

References

1. Rose, C.P., Gaydos, A., Hall, B.S., Roque, A., VanLehn, K.: Overcoming the Knowledge Engineering Bottleneck for Understanding Student Language Input. In: Proceedings of AI in Education. IOS Press, Amsterdam (2003)
2. Chi, M.T., Feltovich, P.J., Glaser, R.: Categorization and representation of physics problems by experts and novices. Cognitive Science 5, 121–152 (1981)
3. Cooke, N.J., Schvaneveldt, R.W.: Effects of computer programming experience on network representations of abstract programming concepts. International Journal of Man-Machine Studies 29, 407–427 (1988)
4. Chase, W.G., Simon, H.A.: Perception in Chess. Cognitive Psychology 4(1), 55–81 (1973)
5. Levchuk, G.M., Gildea, K., Freeman, J., Shebilske, W., Alakke, G., Narakesari, S.: A Benchmarked Experiential System for Training (BEST) for Optimizing Instruction. In: Proceedings of the 12th International Command and Control Research and Technology Symposium, New Port, RI (2007)
6. Levchuk, G., Shebilske, W., Freeman, J.: A model-driven instructional strategy: The benchmarked experiential system for training (BEST) (in review). Submitted to Journal of Mathematical Psychology
7. Stacy, E.W., Walwanis Nelson, M., Colonna-Romano, J.: Using Pedagogical Information to Provide More Effective Scenarios. In: Proceedings of the 2007 Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL (2007)
8. Stacy, E.W., Colonna-Romano: From events to experiences: Focusing scenarios on training effectiveness. In: Walwanis Nelson, M., Lowe, M. (eds.) Advances in the Theory and Practice of Scenario-Based Training–Charting the Road Ahead. Proceedings of the 2008 Conference on Behavior Representation in Modeling and Simulation, Providence, RI (2008)

Conformity out of Diversity: Dynamics of Information Needs and Social Influence of Tags in Exploratory Information Search

Ruogu Kang, Thomas Kannampallil, Jibo He, and Wai-Tat Fu

Applied Cognitive Science Lab
Human Factors Division and Beckman Institute
University of Illinois, 405 N. Mathews Ave.,Urbana, IL 61801 USA
{kang57,tgk2,jibohe2,wfu}@illinois.edu

Abstract. We studied the dynamic effects of information needs and social influence of tags in an exploratory search task. Although initially differences in information needs led to diversity in tag choices, this diversity disappeared as participants collaboratively tagged the same set of resources. Our findings are in general consistent with the notion that people conform to the collective interpretation of contents in an information system. In addition, our results showed that conformity does not only arise out of imitation of behavior, but also from the same underlying semantic interpretation or knowledge structures of users as they engage in informal collaboration through the social tagging system. Implications for design of social information system are discussed.

Keywords: Exploratory search, tag choice, information needs diversity, semantic interpretation of tags.

1 Introduction

Information seeking activities can be characterized as a form of problem solving (e.g., [6], [8], [14]), in which people are searching for and comprehending information to fulfill their information goals. However, there are often situations in which the information seeker has not yet developed well-defined information goals to guide their search. Instead, the information seeker may have to start with an abstract representation of information needs derived from a broader task context. In these situations, the information seeker has to engage in some forms of *exploratory information search,* through which information goals can be iteratively refined and enriched (e.g., [5], [7]). Recently, researchers have reasoned that the traditional search engines are insufficient for this kind of exploratory search [11]. Instead, many have proposed that the evolving Web 2.0 technologies have great potential for helping people to conduct exploratory information search. However, what is still lacking is a scientific understanding on the interactive cycle of tag-based exploratory search and tag creation in a typical social tagging system. The goal of this paper is to investigate the processes involved in tagging behavior and how it is related to tag-based exploratory search.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 155–164, 2009.

Social tagging systems allow users to annotate, categorize and share information on web resources (links, papers, books, blogs etc.) by assigning tags to them and share the tagged resources with other members of the system. One major reason for the popularity of tagging systems arises from its benefits in supporting exploratory search [13], social navigation [12] and information sharing. Prior research on social tagging systems has primarily focused on characterizing aggregate patterns in tagging behavior (e.g., [3], [4], [9]). However, with a few exceptions [13], there is still a lack of scientific study on how social tags could facilitate exploratory information search.

One important question underlying the success of social tagging system is whether the tags created by the large number of users provide any useful information for others. Indeed, users in any knowledge sharing system may have different underlying motivations to seek and share information, and this may lead to a continuous growth of diversified tags in a single system. For example, a particular user may tag a book based on its content with a tag “Star Trek”, while another user may tag the same book as “To read”, referring to personal intent of this book. Nevertheless, in spite of the perceived unstructuredness, researchers (e.g. [3], [9]) have found long-term stability in tagging behavior. For example, by analyzing a set of data from del.icio.us, Golder & Huberman [7] supported that tag choices are influenced by tags created by other users, even if users may have different information needs when they tag. Sen et.al [16] showed that community influence can directly impact user’s personal tendency in choosing tag vocabulary. A recent study by Rader and Wash [15], however, has raised questions regarding the social influence on tag choices. By analyzing a different set of data from del.icio.us using logistic regression techniques, they found that tag choices could be better explained by personal information needs, which provided evidence against the presumed social influence of tags.

A relevant question regarding this controversy is: what motivates a user’s tag creation (or use) in a social tagging system: personal information needs or social influence? Specifically, we explored how the dynamic effect of social tags and information needs elicits different tag choices among users. Instead of characterizing just the overall aggregate patterns, we specifically focused on users’ information needs to understand how different tag choices emerge. Additionally, we also investigated how people created tags and how tag could facilitate exploratory information search, and how they could be related to each other. To preview our results, we found that tag choices were not only influenced by the information needs of the user, but were also influenced by the semantic interpretation of existing tags. In addition, consistent with results by Millen D.R. et.al [13], we found that browsing was used most often in exploratory information search, and this finding could be explained by theory of *perceived information gain*.

2 Method

2.1 Participants and Platform

Thirty two participants (12 Male, 20 Female; average age =22.6 years, S.D. = 4.5) were recruited from the University of Illinois community. Most Participants rated themselves as moderate computer users with an experience of about 12 years (87.5% browse the web more than once a day).

CiteULike (www.citeulike.com), a research literature sharing website with tagging and search features, was used as our research platform. CiteULike allows users add links to papers and books, and add references from other digital libraries and optionally tag the available content for future reference. We chose CiteUlike mainly for its simplicity of use and the relative ease of creating a library containing book information from external websites. Because CiteULike has all the basic functions of social tagging systems and a large number of users, we believe that our results can be generalized to other systems. User activities including mouse events, URLs, time stamps, and contents of web pages were recorded for further analysis.

2.2 Tasks

The information resource consisted of 150 books that were imported directly from Books@Amazon.com. This "library" of books covered eight categories with approximately equal number of books in each category: Arts & Photography, Business & Investing, Children, Computers & Internet, Cooking, Food & Wine, Health, Mind & Body, Medical, and Self-Help. We designed eight exploratory search tasks to represent different information needs based on the eight categories of books. The search tasks thus provided an abstract representation of information needs for participants to explore the library. For example, participants were asked to find and tag books to recommend for a library in a *retirement community*. The other seven tasks are *Software Company, Local Arts Center, Traveler's Books, Career Center, Rehabilitation Center, Daycare Center, Wellness Center.*

During the experiment, participants were asked to search for books by browsing in the main library, choosing tags from the tag cloud, or using keywords to search. When participants decided to select the book, they were encouraged to create new tags or reuse existing tags for the selected book after reading the description of the book. Participants were instructed to imagine that they were working with a group of other co-workers who were also selecting books for the same or different organizations, and the tags that they created should be useful *not only to themselves but also to others*.

2.3 Analysis

The 32 participants were randomly divided into 4 sessions. In each session, each participant was randomly assigned to one of the eight tasks. In other words, there was exactly one participant assigned to each of the 8 tasks in each session. In addition to controlling the same initial library, we imitated the social environment of tagging system by enabling participants to see all tags created by previous participants.

We compared the tags created by participants on different books across the 4 sessions to investigate how tag choices were influenced by tags created by previous participants. We use Latent semantic analysis (LSA, see [10]) to estimate the semantic relatedness between every set of new tags created and the existing tags in each book selected by the participants. LSA is a statistical technique for extracting and representing the similarity of meaning of words and passages by analysis of large bodies of text. The similarity between resulting vectors for words and contexts has been shown to closely mimic human judgments of meaning similarity. In the current analysis, we performed the LSA calculations through the web site at http://lsa.colorado.edu, using the general reading topic space with 300 factors.

3 Results

Participants selected about the same average number of books across all the sessions ($F(3, 21) = .249$, $p > 0.10$). On average, subjects created 4.76 tags (S.D. = 2.27) for each selected book. As users proceeded through the sessions, fewer tags were created for each selected book ($F(3, 21) = 3.110$, $p < .05$).

3.1 Tag Creation and Tag Choice

The mean number of unique tags assigned to each book decreased across sessions (See Table1), as confirmed by the significant linear downward trend ($F(1, 30) = 3.92$, $p < 0.05$). The decreasing number of unique tags suggests the increasing agreement among participants on the creation of tags to the current set of books.

Table 1. The mean number of unique tags created per book across sessions

Session	1	2	3	4
Number of unique tags	8.0	4.5	3.6	2.9

To understand how different information needs influenced the tag use, we extracted all book-selection episodes from all participants, and calculated the correlation between users' information needs and new tag creation episodes. We believe that tag creation was not solely based on books or tasks, but also on the book "selection" under different tasks, because the "selection" can represent how users interpret the search tasks. Fig. 1 shows how we represent the selections and tags between the task and book space. Each selection in the figure represented the selection of one of the 150 books given that the participant was given one of the eight search tasks. A new-selection between book *i* and task *j* therefore indicated that the participant was the first one to select book *i* under task *j*. Different users may create different selections based on their own information needs and interpretation of the tagged resources.

To measure the relationship between the search tasks and users' motivation to create new tags, we coded each selection as either new or old (1 and 0, respectively) and called this the "new-selection code", and coded each tag creation as either new or old (1 and 0, respectively) and called this the "new-tag code" (See Fig. 1).

Using the coding scheme described above, we calculated the correlation between the new-tag and new-selection codes, which reflected the extent to which the creation of new tags were related to the fact that participants were assigned to a new search task. A high correlation would imply that most of the tag creation occurred when participants were given a new search task. A low correlation would imply that creation of new tags were not related to the search tasks of the participants, in the sense that new tags were created equally often when participants were given the same or different search tasks.

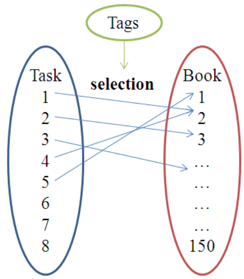


Fig. 1. Users selected books and created tags under different search tasks. Each book-selection episode is represented as an arrow from task to book in the figure. Each selection was associated with a set of tags created by the user. Users could repeat existing selections (select books that had been selected by other participants under the same search tasks before) or create new selections (select books that had not been selected before under the same search tasks). For old selections, tags could be new (not used before for the selected book, regardless of search tasks) or old (reused tags for the selected book).

Table 2. Correlation between the new-tag and new-selection codes in all book-selection episodes across sessions 2 to 4

Session	2	3	4
Correlation between new-tag and new-selection codes	0.62	0.42	0.24

Table 2 shows that the correlation between the new-tag and new-selection codes steadily decreased across sessions. All correlations were significant ($p<0.05$), so was the obvious downward trend ($F(1, 22) = 4.21$, $p < 0.05$), which implied that the creation of new tags was more strongly related to differences in the search tasks (which imposed different information needs) early on than in the later stage. There was no significant difference between the numbers of tags created on the same or different selections. The results were consistent with the notion that ***as the number of tags increased, participants became more likely to agree with the existing tags associated with a book***, even though these tags were associated with distinct selections as defined in Fig. 1.

To further understand whether users tended to conform to existing tags created by others, we calculated the LSA scores for all new tags created, and divided these "new-tag episodes" into whether they were associated with a new- or old-selection (see definitions in Fig. 1). Fig. 2 shows the mean LSA scores for the new-tag episodes. The main effect of new/old selections was not significant, but the interaction between new/old selections and sessions was significant ($F(2, 23) = 3.41$, $p < 0.05$). The LSA scores for episodes on old-selection were not significantly different across sessions, but the LSA scores for episodes on new selection in sessions 3 and 4 were significantly higher than that in sessions 2 and 3 respectively. In other words, the LSA scores stayed approximately at the same level for the old-selection episodes (books selected under the same search tasks) across sessions, but the LSA scores increased significantly across sessions for the new selection (books selected under different

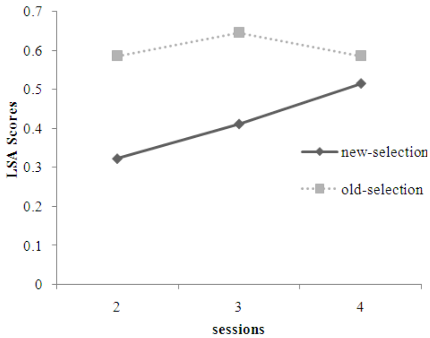


Fig. 2. The mean LSA scores for episodes with new tag creation under new- and old-selection. (In the first session all participants were given new search tasks, so the new-selection codes in all book-selection episodes in the first session were 1 and the correlation is not meaningful)

search tasks) episodes. The LSA score for the new-selection was significantly lower than that for the old-selection in sessions 2 and 3 ($p<0.05$).

Results in Fig. 2 show that initially new tags created under the same search tasks were more semantically related than those created under different search tasks, but as more tags were created, this difference disappeared (see values at session 4). This provides further support to the notion that different search tasks (thus different information needs) have influence on tag choices, presumably because different information needs may prime users to focus on different aspects of a resource or have different interpretation of the information content. However, as the number of tags increased, their "bottom-up" influence on future tag creation increased and eventually outweighed the "top-down" influence from the information needs. This was confirmed by the same level of semantic relatedness for tags under same and different selections in session 4 (see Fig. 2). Over time, the semantic relatedness of tags for a resource increased irrespective of the information need, suggesting that the top-down influence of information need was gradually replaced by the bottom-up influence of existing tags at the semantic level.

3.2 Tag-Based Exploratory Information Search

Exploratory search strategy. In addition to the dynamics of information needs and social influence of tags, we also investigated into the characteristics of exploratory search in social tagging system. First, we extracted the strategies that participants used to select each book to understand how tags influence the use of strategies and how it varied across sessions. There were three major strategies used by the participants: *browsing*, *choosing tags from tag cloud* and *keyword searching*. Most participants selected books by browsing the book titles and tags on the main screen. We called this a *browsing* strategy. The *tag cloud* strategy was when participants clicked on any tag in the cloud and reached a list of books assigned with that tag. The third strategy was *keyword searching*. Participants clicked on "search" and typed in keywords to reach a list of books matching the keywords either in the title or tags.

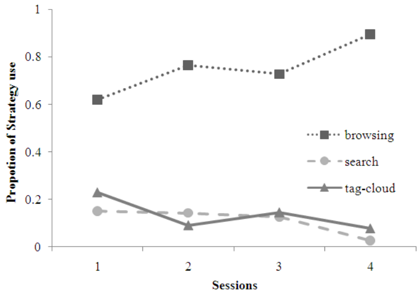


Fig. 3. Proportions of strategy use across sessions

As shown in Fig. 3, the use proportions of the browsing strategy increased but the use of the other two strategies decreased across sessions. Analysis of variance showed that the main effect of strategy was significant ($F (2, 31) = 4.21$, $p < 0.05$), confirming the obviously higher proportion of the browsing strategy than the other two strategies. The strategy × sessions interaction was also significant ($F (6, 31) = 2.43$, $p < .05$), confirming the upward trend for the browsing strategy and the downward trends of the tag-cloud and search strategies.

The results were consistent with the conclusion by Millen [13] that community browsing (topic search) is the most frequently used search strategy in social bookmarking system. One possible reason was that the browsing strategy tends to have higher perceived information gain than the other two strategies. In this task, information gain can be defined in terms of the perceived usefulness of information cues in satisfying the information goal, i.e., the perception that the title words and tags could help participants to judge whether a book is relevant or not. Indeed, previous research has shown that selection of information search strategies is sensitive to the moment-to-moment measure of *information gain per unit time cost* ([1], [2], [8], [14]). One important implication from these studies was the discovery that selection of information search strategy was more sensitive to the on-going evaluation of information gain than the overall efficiency of the strategy.

In the browsing strategy, the continuous evaluation of the title words and tags could lead to a high perceived information gain, as they provided on-going information for participants to judge the relevance of a book. Thus, the perceived moment-to-moment information gain per unit cost could actually increase as more tags were added to the books. We also believe that the higher perceived information gain for the browsing strategy was particularly prominent in exploratory search, as participants only had a rough idea of what they were looking for (as opposed to a well-defined information task in which, for example, a specific title or keywords were given). Thus, browsing not only allowed participants to find relevant books, but also allowed them to *refine and enrich their information goals* as specified in the search task. As a result, the perceived information gain was likely to be higher for the browsing strategies than the other two strategies.

Failed search episodes. To further test our assumption that tags will facilitate exploratory search, we counted the number of events when participants clicked on a book title, read its content description and decided *not* to select the book across sessions. We called each of these episodes a failed-search. If the tags associated with each book indeed provided more useful information for participants to judge the relevance of the book, the number of failed-search episodes should have decreased across sessions as more tags were assigned. Indeed, we found that the number of failed-search episodes decreased steadily across sessions as more tags were added to the library (table. 3), as confirmed by the significant linear downward trend ($F(1,30)=4.93, p<0.05$).

Table 3. Total number of failed-search episode for all participants across sessions

Session	1	2	3	4
Total number of failed search	57	46	32	29

This pattern[1] was consistent with the idea that new tags provided higher information gain for participants as they browsed through the books and the new tags did help them to judge the whether the book should be selected or not without looking at the detailed description of the book. In other words, *participants' judgment on the relevance of books based on these tags actually improved as more tags were added.*

4 Discussion

The current findings provide a novel explanation of tag creation from the semantic level and support the assumption that tags can facilitate exploratory search. Through the collaborative tagging effort by multiple users, the collective interpretation of information content becomes a more important factor that influences tagging behavior than differences in information needs, which is represented by the increasing semantic relatedness between tags. Not only did social tags influence how likely users may come up with new tags to describe the same information resource (Table 1), but when they did decide to create new tags, these new tags were closer in meaning to existing tags (Fig. 2). Researchers have argued that one reason why proportions of tag use tend to stabilize is because people tend to imitate others behavior [9]. Our results provide direct empirical evidence supporting this idea; but in addition to that, our results also highlight the dynamic interaction between personal information needs and social tagging behavior across time, which provides a more in-depth explanation to the stabilization pattern in large scale tagging systems.

Although initially different information needs led to creation of semantically different tags, as more tags were assigned to the resources across a session, new tags created became semantically more similar to existing tags (note that across sessions participants were given the exact same set of different information needs). In fact, in the last session, the level of semantic relatedness between new and existing tags created by participants with different information needs was about the same as those created by participants with the same information needs. The increase in semantic relatedness between new and existing tags across sessions is a novel finding that

[1] The count was pooled across all strategies as they involved the clicking on the book title before they could select the book.

provides strong empirical support for the social nature of tags: As tags accumulated from the collaborative efforts from participants with different information needs, these tags presumably provided increasing depth and breadth to the description of the books. In fact, we believe that a major function of social tags is to provide semantic cues to the content of a resource, so that other users can utilize these semantic cues to estimate the relevance of a particular information resource with respect to their own information need. Thus, it is no surprise that users are conforming to certain underlying semantic structures through repeated interactions with the social tagging system. Our next set of studies will focus on extending the current findings to other "semantically rich" domains (e.g., biologists tagging research articles) to further test this idea.

Lastly, our results have significant implications for the design of next generation social tagging support tools. We established that social tags can facilitate information search because the failed search number significantly decreased across sessions and users chose the search strategy based on perceived information gain. Therefore, incorporating the semantic relatedness between existing tags would provide users a better representation for developing an understanding about current tags and help them to create new tags that would provide more informational value for the user and improve the effectiveness of information search. Given that our results show that the tagging process is sensitive to user's interpretation of existing tags, we speculate that different knowledge backgrounds may influence users' tagging and searching behavior by influencing their interpretation process. Our future study will look into how different knowledge structure may influence users' tagging behavior.

References

1. Blackmon, M.H., Kitajima, M., et al.: Tool for accurately predicting website navigation problems, non-problems, problem severity, and effectiveness of repairs. In: Proceedings of CHI 2005, Portland, OR (2005)
2. Brumby, D.P., Howes, A.: Strategies for Guiding Interactive Search: An Empirical Investigation into the Consequences of Label Relevance for Assessment and Selection. Human-Computer Interaction 23(1), 1–46 (2008)
3. Cattuto, C., Loreto, V., et al.: Semiotic dynamics and collaborative tagging. PNAS 104(5), 1461–1464 (2007)
4. Farooq, U., Kannampallil, T.G., et al.: Evaluating tagging behavior in social bookmarking systems: metrics and design heuristics. In: Proceedings of the 2007 international ACM conference on Supporting group work, Sanibel Island, Florida, USA, pp. 351–360 (2007)
5. Fu, W.-T.: The microstructures of social tagging: a rational model. In: Proceedings of CSCW 2008, San Diego, California, USA, pp. 229–238 (2008)
6. Fu, W.-T., Gray, W.D.: Suboptimal tradeoffs in information seeking. Cognitive Psychology 52(3), 195–242 (2006)
7. Fu, W.-T., Kannampallil, T.: Harnessing Web 2.0 for context-aware learning: The impact of social tagging system on knowledge adaption. Educational social software for context aware learning: Collaborative methods and human interaction. In: Lambropoulos, N., Margarida, R. (eds.) IGI Global (in press)
8. Fu, W.-T., Pirolli, P.: SNIF-ACT: A cognitive model of user navigation on the World Wide Web. Human-Computer Interaction 22(4), 355–412 (2007)
9. Golder, S.A., Huberman, B.A.: Usage patterns of collaborative tagging systems. J. Inf. Sci. 32(2), 198–208 (2006)

10. Landauer, T.K., Dumais, S.T.: A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. Psychological Review 104(2), 211–240 (1997)
11. Marchionini, G.: Exploratory search: from finding to understanding. Commun. ACM 49(4), 41–46 (2006)
12. Millen, D.R., Feinberg, J.: Using social tagging to improve social navigation. In: Workshop on the Social Navigation and Community Based Adaptation Technologies, Dublin, Ireland (2006)
13. Millen, D.R., Yang, M., et al.: Social bookmarking and exploratory search. In: ECSCW 2007, Limerick, Ireland (2007)
14. Pirolli, P., Card, S.K.: Information foraging. Psychological Review 106, 643–675 (1999)
15. Rader, E.J., Wash, R.: Influences on tag choices in del.icio.us. In: Proceedings of CSCW 2008, pp. 239–248. ACM Press, San Diego (2008)
16. Sen, S., Lam, S.K., et al.: Tagging, communities, vocabulary, evolution. In: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work, pp. 181–190. ACM Press, Banff (2006)

Trail Patterns in Social Tagging Systems: Role of Tags as Digital Pheromones

Thomas George Kannampallil and Wai-Tat Fu

Applied Cognitive Science Lab
Beckman Institute and Human Factors Division,
University of Illinois at Urbana-Champaign,
405 North Mathews Avenue, Urbana, IL 61801
{tgk2,wfu}@illinois.edu

Abstract. The popularity of social information systems has been driven by their ability to help users manage, organize and share online resources. Though the research exploring the use of tags is relatively new, two things are widely acknowledged in the research community: (a) tags act as a medium for social collaboration, navigation and browsing and (b) an overall stable equilibrium exists among tag patterns due to the social nature of the tagging process. But there is very little agreement on what causes these stable patterns. In this paper, we take an evolutionary perspective to understand the process of tagging to investigate whether tags act as "way finders" or digital pheromones in social tagging systems. We investigate the existence of tag trails based on a semantic similarity measure among existing tags. We found that over 50% of the resources we evaluated exhibited strong trail patterns. The implications of these patterns for the design and management of social tagging systems is discussed.

Keywords: Social Tagging Systems (STS), Stigmergy, Pheromones, Web 2.0.

1 Introduction

With the widespread popularity of Web 2.0, social tagging is a common feature in several web-based systems. Social tagging systems (STS) provide the flexibility to its users for annotating, organizing and categorizing their information content on the web using tags. Tags help in supporting online search, navigation, managing content and sharing with other users who have similar interests. Examples of social systems include del.icio.us (http://del.icio.us) (URLs), Flickr (http://flickr.com) (for photos), CiteULike (www.citeulike.org) (for academic papers and books). Due to its widespread popularity and use, tagging systems contain hundreds of thousands of users, tags and resources (URLs, photos, books etc.). Users of STS are free to create tags at their free will and in most cases user activities are not moderated by administrators. In this environment, one would expect relative chaos and unstructuredness in tagging patterns. But, researchers have established that stable tagging patterns arise in social tagging systems (e.g., [3, 8]).

The stability of tagging patterns has been attributed to a variety of reasons: social imitation [8], semantic organization and clustering [11], information theory based

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 165–174, 2009.

models [4], preferential attachment [3, 10], personal preferences [14], recency of tag use [3] and a rational model [6]. The most widely accepted idea on overall tag stability is *social imitation*. Researchers [8] have argued that tag equilibrium is achieved through users' direct imitation of tags created by other users. They suggest that users directly use the tags that are previously created. The underlying assumption on the occurrence of social imitation is that tags act as a medium of *direct* coordination activity.

We believe that there are more subtle nuances by which users appropriate tags (e.g., by creating tag synonyms, semantically similar tags, or some personalized form of tags). Thus a direct imitation model may not completely explain tag pattern equilibrium. We investigate the *effect of the social medium on the coordination practices* during tagging and the role of indirect *social coordination* in the creation of stable tagging patterns. We use the principles of stigmergy, a mechanism for explaining the indirect coordination between agents (or humans), to investigate the causes for stable patterns of tagging behavior.

Stigmergy is based on the idea that physical traces of work left by others in a medium act as the basis for future coordination activities. The idea of stigmergy was developed by Grassé [9] to describe the emergence of collective coordination activities of social insects. The concept was initially used to explain the *coordination paradox* in group activities: i.e., looking at a group of social insects (Grassé looked at nest building activities of termites), it would seem that they are cooperating in an organized manner, but looking at an individual would present the picture of independent work and not being involved in the collective activity. The explanation based on stigmergy for coordination paradox is that the collective interaction is *indirect* [15]. In other words, the agents affect the behavior of other agents through indirect communication using social artifacts in the physical environment. For example, in the case of termites, nest building material; in the case of ants, ant trails are supported by pheromones. A detailed review of stigmergy can be found in Theraulaz and Bonabeau [15].

One of the most popular examples of stigmergy is the food tracing behavior in ant-colonies. Ants in the real world wander randomly between the food source and their colony. The (initial) ants leave *pheromones* in their trail as they move from the food source to the colony (and vice versa). Other ants are more likely to follow the pheromone trail rather than a random trail, thereby reinforcing a previously existing trail. If the path to the food source is long, the pheromone trail evaporates over time. Alternatively, ants choosing shorter paths will have their pheromone trails reinforced by consistent ant-traffic and shorter path length. As a result of the pheromone evaporation in longer paths, less preferred longer routes are no longer followed by ants. But when a (random) ant finds a shorter path, other ants are likely to follow that path resulting in an overall positive feedback along that path [5]. In this example, ants coordinate their action through the indirect interaction with the physical medium and pheromones act as the medium for their coordination activities.

There are many parallels between a social tagging environment and the ant-colony described above. Users in social tagging systems are driven by their *local* goals that are driven primarily by their information needs. The *global* behavior of the users are emergent and occurs as they use the social tagging system. The global behavior is spurred by the instinctive response to traces in the medium (in this case, tags). The trail strength is developed as more users add tags that are similar to the existing tags.

Table 1. A comparison between an ant food searching pattern and social tagging

	Ant Colony	Social Tagging
Local behavior	Ants searching for food	Users searching STS for their information needs
Coordinating medium	Pheromones left on trails by previous ants	Tags that are added by prior users of a resource; tags act as "digital pheromones"
Trail Strength	Shorter trails have more ant traffic leading to trail strengthening	Addition of semantically similar (or same) tags leads to a stronger strength for that tag as a descriptor for the resource
Global behavior	Ants have a coordinated shortest path trail to the food source	Coordinated overall global stable pattern across the STS

This leads to later users perceiving the resource in a particular way. The comparison between an ant colony and social tagging is shown in table 1.

The concept of stigmergy is extremely relevant in the case of social tagging activities. We believe that social tagging systems present a medium where collective action is based on the distributed cognitive activities of a set of users. The indirect coordination practices can be considered as the interconnecting glue for distributed cognitive system, with users, resources and tags, creating a balance between individual action and social phenomena. But we know very little about how these users coordinate their tagging behavior without any direct communication. We believe that the coordination activities are *indirect* and are an effect of the tagging environment. We hypothesize that similar to the ant-colony environment, tags act as digital pheromones, creating trails for users. The digital trails are strengthened by the addition of semantically similar tags by other users. Specifically, we investigate the following research questions: (a) Do tags act as "digital pheromones" to support indirect coordination? (b) How do these trails affect the overall equilibrium in tagging patterns?

We define a *trail strength index* to evaluate the strength of a trail based on the semantic similarity between tags. Let us consider the following scenario: A user adds a tag "network" to a paper on "Facebook use". The tag "network" acts as a digital trace in the medium for future users for the paper. The addition of a new tag (or modification of an existing) that is semantically similar to the existing tag would mean that the existing trail for the tag "network" is strengthened. In other words, a new user indirectly has a general degree of collective agreement on a resource. As more tags are added, the presence of semantically similar tags strengthens the trail in a certain direction. Thus, if more tags similar to the tag "network" are added, the paper becomes perceived by future users as a paper related to networks based on its tags. Addition of tags that are not semantically related to network (e.g., a tag "food") would lead to lesser strength to the tag trail. Using data from the popular scholarly social tagging system CiteULike, we investigate the development of trails in tagging networks. We found that over 50% of the resources exhibited the strong trails, where a digital trail was created with semantically similar tags. About 23% of the resources had weak trails phenomena, where there was little agreement among the taggers (nor were semantically similar tags) by later users. We describe the implications of strong and weak trails for growth of stable patterns and discuss its importance for the design and management of social tagging systems.

2 Process of Social Tagging

There are three main components for any tagging system: users, tags created by the users and resources (URL, books, pictures, movies etc.) for which tags are assigned.

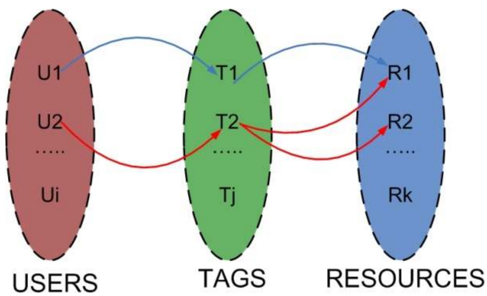


Fig. 1. Process of tagging: User 1 (U1) has one tag (T1) applied to one resource (R1) (blue arrow) User 2 (U2) has one tag (T2) applied to two resources (R1 & R2) (red arrow)

The primary difference between different tagging systems is the type of resource. Resources can be different depending on the specific purpose of the social tagging system. For CiteULike (www.citeulike.com), a resource is an academic paper or book, while for del.icio.us it is a web site URL. Other media such as photos and videos have also been resources for social tagging systems. As explained earlier, a tagging system can be represented as a tuple: users, tags and resources (see U, T, and R in Figure 1). For example, consider that a user U1 applies tag T1 to a resource (R1) (see blue arrows) and user U2 applies the tag T2 to two resources R1 and R2 (see red arrows). Each of these assignments is called a tag application. Thus there are three tag applications in this system (U1, T1, R1; U2, T2, R1; U2, T2, R2). This is the widely accepted conceptual model of a social tagging system. Large social tagging systems contain hundreds of thousands of such tag applications resulting in significant interaction patterns among tags such as re-use, growth and informational value.

3 Method

In this section, the data source and the related analysis methods that were used for this study are described.

3.1 CiteULike Tagging Data

The data for this study consisted of tagging data from the popular social tagging system, CiteULike. CiteULike is an online social bookmarking service supporting the storage, sharing and organization of information on research papers. It is primarily used by researchers. CiteULike users can link papers and import references from other scholarly digital libraries. User can also add their favorite papers to their collection and assign tags to them. Our data consisted of tagging data from CiteULike over a 5-month period from November, 2004 to March, 2005. In total, there were 65,347 tag applications by 1205 users and 12067 total unique tags. The source of the data was publicly posted logs from www.citeulike.org. The data was processed to extract the user-tag-resource relationship during this time period.

3.2 Data Analysis

One of the main challenges of large datasets from public web portals is the presence of spam. While an analysis of all the tag applications is certainly useful, the results would likely be spurious due to the presence of significant spam content. As a result, we decided to process a select set of tags and manually ascertain its quality to avoid this problem. The CiteULike tagging data for the 5 month period was organized in a database. We then randomly selected 100 resources (books or papers) that had at least 5 tags and was tagged at least by 5 different users. This was done for two purposes. First, we needed to have a clear *time-sequence of tagging events* (hence the 5 tag application limit). Second, in order to establish the *effect of indirect coordination practices*, we needed to observe the effect of previous tags on the choice of future tags. The selected set of resources was manually evaluated to remove the spurious tags.

After the first 100 resources were extracted along with the corresponding users, tags and time of tags, Latent Semantic Analysis (LSA) was performed on each set [12]. LSA is used to extract and represent similarity of word meanings by comparison to large corpora of text. The LSA values reflect the general semantic similarity among words. It uses singular value decomposition, a general form of factor analysis, to condense a very large matrix of word-by-context data into a much smaller dimensional representation. LSA, as well as variations of similar statistical language techniques such as information scent [1, 2, 7, 13], had been successfully applied to explain how users interpret the relevance of link text on web pages (e.g., [1, 2]). Prior research shows significant support for the use of LSA as a method for measuring human interpretation of relatedness in text. For our analysis, LSA was performed using the algorithm available through the website at http://lsa.colorado.edu, using the general reading topic space of 300 factors.

Our analysis was conducted in the following manner: the tags created by the first user were selected. Each of these tags is compared with the rest of the tags created for that resource. In other words, the semantic similarity of each initial tag with respect to all the other available tags is computed. The matrix LSA comparison was used for our analysis. For each initial tag, we then computed the average semantic similarity score across all other available tags. We call this *trail strength index (*TSI*)*. The higher the value of TSI, greater the semantic similarity between a tag and its follow up (trailing) tags. Lower values of trail strength indicate that successive users do not use this tag for their tagging processes.

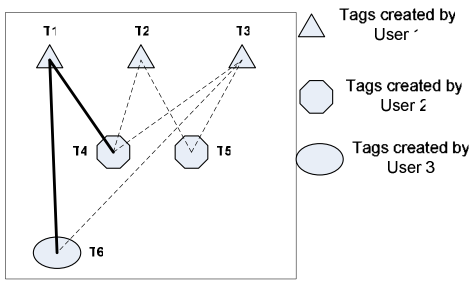


Fig. 2. Evaluating the trail strength index: If a user creates semantically related tags to pre-existing tags, the strength of trail is increased (see dark line between T1 and T4, higher LSA score, therefore greater trail strength). The dotted lines show lower semantic relatedness between tags created by successive users (thus the trail strength is lower across user tagging sessions). The trail strength in this case is for tag T1 (and its semantically related tags T4 and T6) resulting in future user applying tags that are similar to T1 (or semantically similar).

The concept of strong and weak trails is shown in figure 2. As new tags are created, they are compared with prior tags to evaluate their semantic relatedness. If new tags are semantically related to previous tags, then it means users are following a trail left by previous users (hence strengthening the trail). Thus in figure 2, the semantic similarity between T1 (created by user 1) and T4 (created by user 2) is high (shown by the dark line) than with other tags. This trail is strengthened further by the addition of a new tag T6 (user 3) which is again semantically related to T1. Thus the trail of tag T1 is strengthened leading future users to perceive that T1 is the most appropriate tag for that resource. The dotted arrows show tag trails with lesser semantic relatedness. It is also likely that users may follow more than one trail. In summary, *the tags that have a higher TSI score will act as trails for that resource* resulting in future users applying that tag to the resource.

4 Results

In this section we report on the results from our analysis. As explained earlier, the analysis was conducted on 100 randomly selected resources which had more than 5 unique tags assigned to them and also had at least 5 different users tag the resource. We use examples to show two different processes that happen during tagging: strong trails and weak trails.

Of the 100 resources that we used for our analysis 53 of them fell in the strong trails category, while 23 fell in the weak trails category. The rest could not be classified into either category. That is, 76% (76/100) of the resources exhibit some form of trail properties. These trail properties were calculated based on the semantic relatedness between temporally sequenced tags. While strong trails were fairly prominent and easy to identify and explain, weakening of trails was significantly more subtle. In the sections below we present an example of strong and weak trails based on the TSI score.

4.1 Strong Trails

Let us consider an example to explain the strong trails phenomenon. A total of 10 tags were created by 5 users (8 unique tags) for a resource (see table 2 for a sub-set of the tags). Tags were added in the order they are presented in the table (user 1 adds the tags "oscillations" and "synchronization"; then user 2 add the tag "oscillations", etc.). The trail strength index was computed by comparing each of the first two tags to the rest of the tags that were created for the resource over the five month period. Table 2 shows the computed LSA scores (and the tag trail strength index) for the two tags created by the first user.

Table 2. Average latent semantic indices for the first and second tags created by the first user for a resource are computed

	Oscillations (User 1)	Synchronization (User 1)
Oscillations (User 2)	1	0.22
Biochemical (User 3)	0.07	0.08
Bioelectronics (User 3)	0.42	0.19
Networks (User 3)	0.04	0.23
Dynamics (User 4)	0.18	0.12
Oscillations (User 4)	1	0.22
Oscillators (User 4)	0.51	0.23
Trail Strength Index	**0.46**	**0.19**

Two important deductions can be made: (a) the first tag ("oscillations") or a highly semantically related (or similar) tag is used by future users (except User 3, who creates two tags unrelated to previous tags and one tag ("bioelectronics") which is semantically similar to pre-existing tags) leading to a fairly strong trail (avg. score =0.46, in spite of a user creating semantically unrelated tags). Thus a fairly strong "tag-trail" exists for this resource with the tag "oscillations" and (b) the second tag ("synchronization") has low semantic relatedness with other tags and it does not show decay or strengthening across all users. It is difficult to draw anything conclusive about the trail strength of the second tag.

Strong tag trails occurs by the addition of tags that are similar (or same) as the original tag. In the case of this resource, there is a stronger "tag-trail" for the tag "oscillations". Similar to the case of ants where pheromones act as a mechanism for tracing food-trail paths, strong tag trail, in terms of the semantic relatedness between a group of tags, acts as probable guidance for future users. While, it is possible to say that "oscillations" act as a digital trace (or pheromone trail) we cannot establish this unless we investigate its trail over a longer period of time. For this, we extracted all the tags that were created for this resource over the next 24 months. The next eight tags that were created for this resource were the following: oscillation, network, vibration, test, bio, frequency, oscillator, and connections. If the trail strength that has been originally established for the tag "oscillation" is strong, then more users are likely to add a tag that is semantically related to it. For this we computed the LSA scores for the six tags that were created after the 5 month period. The mean LSA score for this set was 0.39 (the TSI score). This means that the trail of the tag "oscillator" was significantly strong and thus it is likely for other users to apply this tag (or a semantically related tag).

4.2 Weak Trails

Weakening of tag trails is the opposite of trail strengthening, i.e., the effects of tags that are created during the early phases of tagging have no effect on the more recent tags. In other words, there is no semantic relatedness between tags created earlier and recent tags. As with strong trails, we use an example to demonstrate the weakening of tag trails. As it can be noted from table 3, the initial tags that are created by user 1 (collaboration and computer-mediated) have very little semantic similarity with future tags. There does not seem to be any semantic relationship between the tags created by the first user and the ensuing tags (the TSI score for tag "collaboration" was 0.03; TSI for tag "computer-mediated" was 0.07). We also did not find any significant semantic relationships between the tags created by the second user and the third user to the later tags.

Table 3. Decaying trail strength index. This resource had more tags but some of these tags did not have LSA documentation (e.g., folksonomy, blog) so they are not shown in this list.

	Collaboration (User 1)	Computer-Mediated (User 1)
Psychology (User 2)	0.01	0.04
Social (User 2)	0.13	0.00
Email (User 2)	0.00	0.00
Technology (User 3)	0.03	0.31
Group (User 4)	0.04	0.01
Psychology (User 4)	0.01	0.04
Communication (User 5)	0.03	0.09
Trail Strength Index	**0.03**	**0.07**

Similar to strong trails, we investigated whether any of the tags created over the following 24 months had semantic similarity to these tags. Based on the set of next 8 tags, we found that there were no pairs of tags that had a tag trail index score greater than 0.1. This means that in the case of weak trails, the tags are semantically "spread" without any clear paths or trails. In such resources, the medium creates a divergence of ideas and concepts. A users' tagging choice becomes less clear in such resources.

5 Discussion

Based on the analysis of data from a social tagging system, we identified the role of tags as digital pheromones, amplifying a significant number of resources with strong tag trails. Resources with strengthened trails exhibit clear "themes" for a resource, while the resources with weak trails had divergent tags without clear "themes" that identify the resource. While it is almost impossible to ascertain the accuracy of the themes in resources with strong trails, it is still a useful benchmark for users who have limited expertise and experience with a topic.

Based on our analysis we found the following: (a) Digital traces of tags are a function of social and semantic imitation. Tag imitation may not be direct, as described by Golder and Huberman [8], but in a more nuanced, semantic manner. (b) Stronger trails do not imply higher informational value for the tags. Tags are generally used in social navigation and search. If a resource has high tag-trail strength, it means that those tags are likely to be general descriptors of the resource and would be highly unlikely to have high informational value during search. (c) There is also the possibility that the tags that act as pheromones (with high tag-trail strength) could not be the best descriptor of the resource (possibly by the lack of or incomplete knowledge of the tag creator). This is a real possibility and could lead to spurious tags for the resource. Conversely, resources that have weak trails are likely to have tags with high informational value (i.e., able to uniquely identify a resource). Lower tag-trail strength is often a function of the nature of the tag (general vs. specific).

Additionally, it is possible to draw some interesting insights on tagging patterns based on the presence of strong and weak tag trails. The resources that have tags with strong trails are exhibit convergence of ideas and are likely to converge to a stable equilibrium much faster. This is because the tags that are added to strengthen the trail (semantically related tags) are concepts and ideas that are closely related to a central theme. One important aspect of trails which cannot be explored with a general dataset such as the one we used is to investigate the effect of expertise on the tag strength. It is likely that users' with lesser knowledge on a topic that create general tags or use their limited knowledge to create tags (resulting in semantically related tags). The creation of semantically related tags by users with lesser expertise or knowledge could also be explained based on the principle of least effort. In other words, adding existing tags (or semantically similar tags) involves less cognitive effort. Users, thereby take the easiest path of adding new tags.

In contrast, resources with weaker trails have tags that are semantically distributed or spread. It is more likely that these tags would converge at a much lower rate and are likely to be very specific (e.g., in our example, "computer-mediated"). These tags have higher informational value for search and navigation purposes.

The identification of strong and weak trails has implications for the design of social tagging systems. First, it is easy for system administrators to identify resources that have strong trails. These can be identified as resources which are less likely to be reached by users and salient tags that are not on the tag-trail can be amplified to help users in their search. Second, resources with weak trails should be amplified with tags in multiple directions to support search and retrieval. This is because these resources would otherwise not be discoverable by a large percentage of users who do not know the specific keywords.

Based on our current study, it is difficult to emphatically establish that stable equilibrium in social tagging systems is caused as a result of indirect coordination mechanisms achieved through the creation of semantically related tag-trails. But the use of semantically tags during tagging is evident (in the case of strong trails). We need to conduct more analyses to explore whether the tags that cause tag trails contribute more towards overall equilibrium than tags that have lower tag-trail strengths. Our results show that indirect coordination using tags is a strong basis for the explanation of social tagging systems as distributed cognitive systems.

References

1. Blackmon, M.H., Kitajima, M., Polson, P.G.: Tool for Accurately Predicting Website Navigation Problems, Non-Problems, Problem Severity, and Effectiveness of Repairs. In: Proceedings of CHI 2005, Portland, OR (2005)
2. Brumby, D.P., Howes, A.: Strategies for Guiding Interactive Search: An Empirical Investigation into the Consequences of Label Relevance for Assessment and Selection. Human-Computer Interaction (2008)
3. Cattuto, C., Loreto, V., Pietronero, L.: Semiotic Dynamics and Collaborative Tagging. Proceedings of National Academy of Sciences 104, 1461–1464 (2007)
4. Chi, E.H., Mytkowicz, T.: Understanding the Efficiency of Social Tagging Systems Using Information Theory. In: Proceedings of the nineteenth ACM conference on Hypertext and hypermedia. ACM, Pittsburgh (2008)
5. Dorigo, M., Maniezzo, V., Colorni, A.: Ant System: Optimization by a Colony of Cooperating Agents. IEEE Transactions on Systems, Man, and Cybernetics–Part B 26(1), 29–41 (1996)
6. Fu, W.-T.: The Microstructures of Social Tagging: A Rational Model. In: Proceedings of the ACM 2008 conference on Computer Supported Cooperative Work. ACM, San Diego (2008)
7. Fu, W.-T., Pirolli, P.: Snif-Act: A Cognitive Model of User Navigation on the World Wide Web. Human-Computer Interaction 22, 355–412 (2007)
8. Golder, S.A., Huberman, B.A.: Usage Patterns of Collaborative Tagging Systems. J. Inf. Sci. 32(2), 198–208 (2006)
9. Grasse, P.P.: La Reconstruction Du Nid Et Les Coordinations Interindividuelles Chez Bellicositermes Natalensis Et Cubitermes Sp. La Th'Eorie De La Stigmergie: Essai D'interpr'Etation Du Comportement Des Termites Constructeurs. Insectes Sociaux 6, 41–81 (1959)
10. Halpin, H., Robu, V., Shepherd, H.: The Complex Dynamics of Collaborative Tagging. In: Proceedings of the 16th international conference on World Wide Web. ACM, Banff (2007)
11. Heymann, P., Garcia-Molina, H.: Can Social Bookmarking Improve Web Search? In: First ACM International Conference on Web Search and Data Mining (WSDM 2008) (2008)
12. Landauer, T.K., Dumais, S.T.: A Solution to Plato's Problem: The Latent Semantic Analysis Theory of the Acquisition, Induction, and Representation of Knowledge. Psychological Review 104, 211–240 (1997)
13. Pirolli, P., Card, S.K.: Information Foraging. Psychological Review 106, 643–675 (1999)
14. Rader, E., Wash, R.: Influences on Tag Choices in Del.Icio.Us. In: Proceedings of the ACM 2008 conference on Computer Supported Cooperative Work. ACM, San Diego (2008)
15. Theraulaz, G., Bonabeau, E.: A Brief History of Stigmergy. Artificial Life 5, 97–116 (1999)

Real-Time Emotional State Estimator for Adaptive Virtual Reality Stimulation

Davor Kukolja, Siniša Popović, Branimir Dropuljić, Marko Horvat, and Krešimir Ćosić

University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, Croatia
{Davor.Kukolja,Sinisa.Popovic,Branimir.Dropuljic,Marko.Horvat, Kresimir.Cosic}@esa.fer.hr

Abstract. The paper presents design and evaluation of emotional state estimator based on artificial neural networks for physiology-driven adaptive virtual reality (VR) stimulation. Real-time emotional state estimation from physiological signals enables adapting the stimulations to the emotional response of each individual. Estimation is first evaluated on artificial subjects, which are convenient during software development and testing of physiology-driven adaptive VR stimulation. Artificial subjects are implemented in the form of parameterized skin conductance and heart rate generators that respond to emotional inputs. Emotional inputs are a temporal sequence of valence/arousal annotations, which quantitatively express emotion along unpleasant-pleasant and calm-aroused axes. Preliminary evaluation of emotional state estimation is also performed with a limited set of humans. Human physiological signals are acquired during simultaneous presentation of static pictures and sounds from valence/arousal-annotated International Affective Picture System and International Affective Digitized Sounds databases.

Keywords: Real-Time Emotional State Estimator, Adaptive Virtual Reality Stimulation, Artificial Neural Network, Stimuli Generation, Physiological Measurements.

1 Introduction

Research described in this paper is a part of ongoing efforts to design and develop the physiology-driven adaptive stimulation for VR exposure therapy (VRET) [1,2]. In VRET, a treatment method for various anxiety disorders, the therapist (*the supervisor*) operates a user interface to deliver gradually to the patient (*the subject*) the virtual stimuli of anxiety-provoking situations [3,4,5]. Physiology-driven adaptive VR stimulation attempts to optimize and customize the therapy by relieving the supervisor of repetitive interface manipulation and monitoring of the subject's physiology. However, it may also be useful in a broader range of human-computer interaction applications.

Physiology-driven adaptive VR stimulation includes [1]: time-synchronized stimuli generation, acquisition of the subject's physiological response, subject's emotional

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 175–184, 2009.

state estimation, and closed-loop control that leads to subsequent generation of new stimuli. Control signals may specify semantics, emotional properties, and media form of the stimuli. The stimuli are presented in various media forms, like static pictures, sounds and synthetic virtual stimuli combined with real-life video clips. Emotional state estimation is based on the dimensional model of emotions organized along the axes of valence (unpleasant-pleasant) and arousal (calm-aroused) [6]. Mappings to other emotion representations may be added if necessary. For example, representation of relevant emotional states in VRET may be more coarse-grained [2]: non-aroused, aroused and overly aroused.

The paper is focused on the real-time Emotional State Estimator component of the physiology-driven adaptive VR stimulation. This component is crucial for adaptation of VR stimulations to the emotional response of each individual. The remainder of the paper lays out design and preliminary evaluation of the real-time Emotional State Estimator based on artificial neural networks, also describing accompanying stimuli generation and physiological measurements.

2 Stimuli Generation

Within the physiology-driven adaptive VR stimulation, the Stimuli Generator is responsible for finding the best-matching stimuli with respect to the semantics, emotional properties and media form specified by the control signals. In this process, it is important that the signals result in emotionally and semantically aligned stimuli, which are individually conformed to a specific subject's mental state.

The Stimuli Generator uses emotionally and semantically annotated stimuli databases. International Affective Picture System (IAPS) [7] and International Affective Digitized Sounds (IADS) [8] are such publicly available databases of static pictures and sounds. Valence and arousal emotional annotations of IAPS and IADS stimuli are decimal numbers in the range from 1.00 through 9.00, representing maximum unpleasantness through maximum pleasantness, and maximum calmness through maximum arousal, respectively. These databases also use free-text keywords, or tags, to describe the semantics of stimuli. However, the keywords are semantically scattered, taxonomically disordered, and subsequently cumbersome for information extraction.

Stimuli generation plans to introduce ontology-based tagging in the existing emotionally and semantically annotated databases, in order to achieve more informative descriptions of stimuli and more efficient extraction of context knowledge [9]. This work builds on the current Stimuli Generator that generates IAPS and IADS stimuli [10]. Media forms supported by the Stimuli Generator are also being extended to virtual stimuli in the context of VRET [11].

3 Physiological Measurements

Two physiological measurement approaches are used during the research and development of the physiology-driven adaptive VR stimulation. One approach involves artificial subjects that computationally generate physiological signals, and the other acquisition of physiological signals from real human subjects. Artificial subjects approach is convenient for software development of the physiology-driven adaptive VR stimulation. It allows generating numerous subjects and measuring their physiological reactions to a large number of stimuli. This facilitates development of the real-time Emotional State Estimator without requiring time-consuming and sophisticated emotion elicitation and physiological measurements on real humans. Approach with human subjects is relevant for the ultimate goal of estimating human emotional state, as artificial subjects provide only crude approximations of human physiological response to stimuli.

Artificial subjects consist of parameterized skin conductance (SC) and heart rate (HR) generators that accept valence/arousal of the generated stimuli (Fig. 1). Generated SC and HR signals exhibit simpler and more deterministic behavior than human physiological signals. However, some principles from the literature on relationship between physiology and valence/arousal have been incorporated, including initial post-stimulus HR deceleration dependent mostly on valence [12], HR acceleration affected by increase of arousal [12], and SC increase with increase of arousal [12]. Approximate parameter values of SC response have been found in [13] and the basic SC response model has been adopted from [14]. HR generator is a modification of the open-source generator ECGSYN [15] with parameter adjustments based on valence/arousal inputs. Description of the underlying modeling used in ECGSYN generator can be found in [16].

Human physiological signals can be acquired by a variety of multi-channel physiological acquisition systems. Acquisition system used in the paper is BIOPAC MP 150. This system is synchronized with SuperLab stimulus presentation system, which presents emotion eliciting stimuli to the human subjects. In order to collect the same signals for humans as for the artificial subjects, acquired physiological signals include

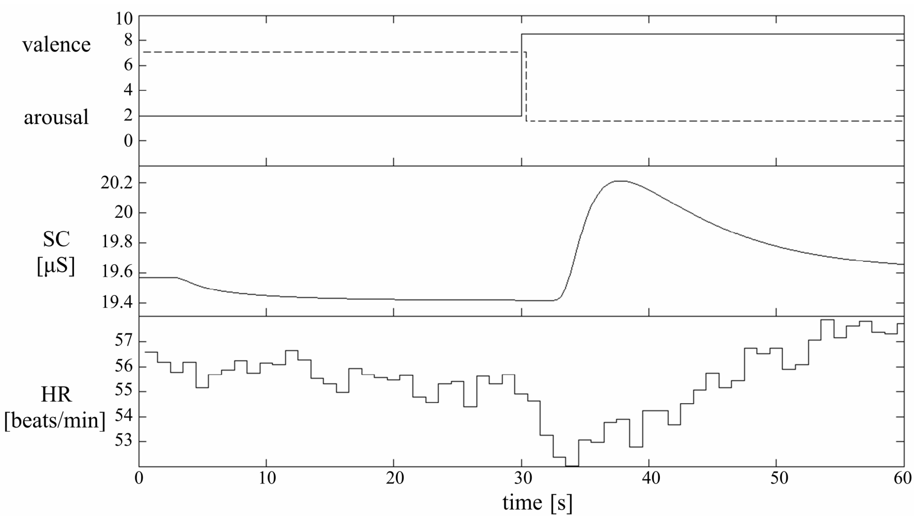


Fig. 1. An example of generated SC and HR reaction to a “step” input, a change from a pleasant relaxing stimulus to an unpleasant arousing stimulus

SC and ECG signal. RR interval algorithm provided with the acquisition system is used in computing the HR signal from the ECG. In-house physiological acquisition system has also been developed, which is more convenient for physiology-driven adaptive VR stimulation.

4 Real-Time Estimation Concept Design

Real-time Emotional State Estimator in physiology-driven adaptive VR stimulation estimates the subject's emotional state repeatedly as physiological signals are acquired (Fig. 2). Subject's emotional state is changed in response to the presented stimuli and is generally affected by other influences, internal or external to the subject. The paper investigates emotional state estimation with sequential delivery of stimuli. *Stimuli sequence* is represented as $s^1, s^2, ..., s^m$, where stimulus s^i has an associated valence/arousal annotation (v^i, a^i) in the stimuli database. Corresponding *stimuli durations* in seconds are $d^1, d^2, ..., d^m$, and *stimuli onset times* $t^1 := 0, t^2, ..., t^m$, with $t^i := t^{i-1}+d^{i-1}$, for $i = 2, ..., m$.

Frequency of outputting the estimated valence/arousal is a predetermined fixed number of outputs per second, called the estimator *framerate* f_e. Acceptable real-time framerate in VRET is 1 Hz. Estimator *frame* is a time interval, in seconds, that extends from the beginning to the end of computation eventually producing one estimator output. *Frame duration* is $T_e := 1/f_e$.

For each frame, valence/arousal emotional state is estimated from physiological samples acquired during a time interval that starts a number of seconds in past and ends at the beginning of the frame. This interval is the *frame window* w_j^i, indexed in the same manner as the beginning of the frame. Physiological samples from the frame window are used in computing the *features* required for valence/arousal estimation. Generally, frame windows can have equal fixed *length*, in seconds, or the length may

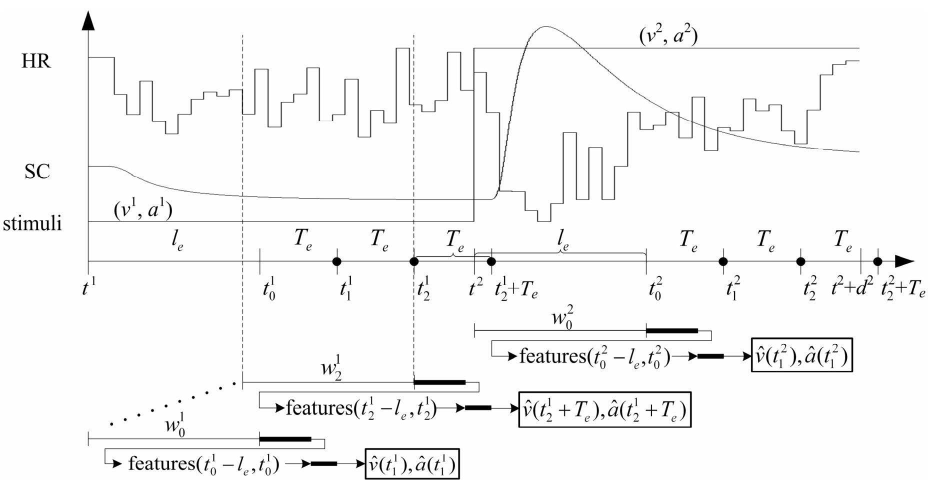


Fig. 2. Real-time emotional state estimation

vary from frame to frame, e.g. to adjust to the dynamics of the features. Frame windows of fixed length l_e are applied, since the logic that may adjust window length to the feature dynamics is beyond the scope of the paper.

The Emotional State Estimator is *disabled* (does not estimate valence/arousal) whenever the frame window includes onset time of a stimulus; otherwise, it is *enabled*. Frame windows that overlap with a change from one stimulus to the next are avoided in emotional state estimation, as they introduce conceptual difficulties in defining the subject's emotional state. Illustration is given in Fig. 2, where no frame window overlaps with stimulus onset time t^2 and, therefore, no black-point markers of estimator output appear on time axis between $t_2^1 + T_e$ and t_1^2. Ability to disable and enable the estimator may rest in some higher-level logic with information regarding the frame windows and stimuli onset times.

5 Emotional State Estimator Based on Artificial Neural Networks

There is a variety of research works that address emotion recognition based on physiology. Numerous methods for emotion recognition have been used, like k-nearest neighbor [17,18,19], discriminant analysis [17,18,20], support vector machine [19, 21], artificial neural network (ANN) [17,18,22,23], Bayesian classification methods [19] or regression tree [19]. In the surveyed articles, majority of the focus is on discrete emotion recognition, like pleasure, sadness, fear, anger etc. One encountered article [22], investigates valence and arousal estimation.

Complementing the previous research, the Emotional State Estimator presented in this paper gives unquantized real-time estimation of valence and arousal based on extracted physiological features. This is important in order to implement the closed loop of adaptive control presented in [1,2].

Estimator design is based on ANNs, due to their ability to model complex relationships between inputs and outputs. Unlike previous similar research [22], several ANN designs are tested:

1. One multi-input multi-output feedforward ANN with two output nodes, for valence and arousal,
2. Two separate feedforward ANNs for valence and arousal, each with a single output node,
3. One multi-input multi-output ANN with two output nodes, for valence and arousal, with feedback from output to input,
4. Two separate ANNs for valence and arousal, with feedback from output to input.

All designs have one hidden layer with 10 neurons, tansig activation function for hidden layer and purelin for output layer.

Training samples for the ANN are obtained by joining the features extracted from the physiological signals with the emotional annotations of the presented stimuli. Training samples are generated according to the concept of real-time estimation in the previous section; one training sample is obtained from each frame window. SC and HR signals of each subject are firstly divided by their respective mean at baseline, obtained during the initial neutral stimulus, for robustness to inter-subject baseline variation. After this transformation, two preliminary feature sets are extracted from both signals. Common features in both sets are mean, standard deviation and slope; the first set (FS1) further includes minimum and maximum, and the second (FS2) includes difference of maximum to minimum and difference of means between the current and the previous frame window. Every feature is normalized to [–1, 1] range, across all subjects used in training, by linearly mapping the minimum value of the feature to –1 and maximum value to 1. Stimuli emotional annotations in the training samples are normalized by linear mapping of their original range [1.00, 9.00] to [–1, 1]. The supervised training is performed with Levenberg-Marquardt learning algorithm, including early stopping for enhanced generalization.

Using emotional annotations of the presented stimuli for estimator training, as in this paper, is one of the possible training approaches (e.g. [22]). Another approach includes training the estimator with the subject's self-reported emotional state for each presented stimulus (e.g. [19]). Selected approach avoids the effect that human subjects' mental construction of self-report may have on physiology. Downside of the selected approach, however, is its inability to incorporate inter-subject differences in emotional experiences of the same stimulus.

6 Preliminary Estimator Evaluation

Accuracy of the estimated valence/arousal is evaluated against the valence/arousal values from the stimuli emotional annotations. Evaluation results are reported separately for valence and arousal, in terms of mean absolute error (mean AE) and maximum absolute error (maximum AE) over all subjects and stimuli sequences in the testing set. Testing set is kept separate from both the training set and the validation set, which is used for early stopping of the training process. Each subject's collected data are exclusively assigned either to the training, validation or testing set.

Evaluation is performed separately for artificial and human subjects, thus assessing inter-subject generalization in both cases. This differs from the protocol in reference [22], which performs valence/arousal estimation with a single human subject. Results are reported only for the best ANN design and feature set, which minimize the Euclidean norm of valence and arousal mean AEs.

Even though evaluation is performed offline in MATLAB, it is carried out in a manner suitable for real-time implementation. Estimator frame duration is set to 1 second, in order to achieve 1 Hz real-time framerate. Normalization of each feature during evaluation relies only on the minimum and maximum values computed during the training. Frame window length is 5 seconds.

6.1 Artificial Subjects

Artificial subjects process the sequences in which stimuli are represented as valence/arousal pairs with associated onset times. Two analyzed cases include evaluation on a variety of stimuli sequences, and evaluation on the sequence that is also used with human subjects. Duration of each stimulus in any sequence is set to 30 seconds.

In the first case (called "stimuli 1"), 10 artificial subjects are exposed to 10 stimuli sequences, each sequence having 21 stimuli. The first stimulus in each sequence has valence and arousal values of 5.00, and represents the ideal neutral stimulus for measurement of the subject's baseline values. Other valence/arousal pairs are randomly selected from IAPS. Six artificial subjects are used for training of the estimator, two subjects for validation and two for testing. The best results are presented in Table 1, achieved by ANN design 4 and feature set FS2.

Table 1. Estimation errors for the best ANN designs and feature sets with artificial subjects (rounded to two decimal places)

		mean AE	maximum AE
stimuli 1	valence	0.64	4.37
	arousal	0.34	3.58
stimuli 2	valence	0.46	1.77
	arousal	0.32	1.42

In the second case ("stimuli 2"), valence/arousal pairs and onset times from the human stimuli sequence are used as inputs for the artificial subjects. Human stimuli sequence starts with a teal background, to establish the baseline physiology of the subject. Teal color has been chosen as the intermediate hue with the best ratio of elicited positive to negative emotions in a study with college students [24]. The sequence proceeds with 8 pairs of IAPS pictures and looping IADS sounds of matching onset and duration, having as similar as possible semantics and valence/arousal values. The first 2 stimuli are neutral, followed by a group of 3 pleasant relaxing stimuli and, finally, a group of 3 unpleasant highly arousing stimuli. Averages and standard deviations of valence and arousal within each group of stimuli, rounded to two decimal places, are (5.69±0.33, 4.50±0.48), (7.46±0.51, 3.27±0.05), (1.92±0.30, 6.90±0.31), respectively. Four randomly selected artificial subjects are used in training, one in validation and one in testing, to match the numbers of human subjects whose results are given in the next section. The best results are again achieved by ANN design 4 and feature set FS2 (Table 1).

With many random stimuli, arousal estimation is superior to valence estimation, as exemplified by the nearly twice lower arousal than valence mean AE in "stimuli 1". Generated physiological signals indeed provide more information regarding arousal than valence, as valence affects only initial HR deceleration after the stimulus onset. Improvements in valence mean AE and maximum AEs from "stimuli 1" to "stimuli 2", may reflect heavily polarized structure of the human stimuli sequence, i.e. strong negative correlation between valence and arousal.

6.2 Human Subjects

Six male students, 24.0±0.9 years of age, participate in the experiment. Stimuli are delivered via eMagin Z800 3DVisor head mounted display (HMD) with earphones, in order to help the subjects focus on the stimuli. Experiment is conducted in a dim air-conditioned technical laboratory with ambient temperature of 23–24°C. After reading of instructions to the subject, applying the electrodes and the HMD, the subject is left

Table 2. Estimation errors for the best ANN design and feature set with human subjects (rounded to two decimal places)

		mean AE	maximum AE
stimuli 2	valence	1.33	5.09
	arousal	0.74	3.24

to rest for about three minutes, and the stimuli presentation and physiological acquisition are started. Four subjects are used in the estimator training, one in validation and one in testing. The best results are in Table 2, achieved by ANN design 3 and feature set FS1.

Mean and maximum AEs exhibit twofold to threefold increase from artificial to human subjects (cf. "stimuli 2" in Table 1). Therefore, generalization of estimation to the testing set, which includes a subject not encountered during training, is more problematic for humans. This underscores complexity of human individuality relative to SC and HR generators, mentioned in section 3.

Evaluation of different ANN designs and feature sets indicates that they may also affect generalization of estimation to previously unseen subjects. Moreover, feature set FS2 seems better suited for artificial subjects, probably because artificial physiological signals exhibit stronger tendency than human signals to settle in a steady state after each stimulus. Reasons behind the differences in the best ANN design for artificial versus human subjects remain to be elucidated by further experiments.

In order to assess if any generalization might have happened with human subjects, valence and arousal mean AEs from Table 2 are benchmarked against the expected mean AEs of two simplified estimators. The first one performs unbiased constant estimation that always returns the mean values of valence and arousal for the stimuli sequence; the human stimuli sequence has both mean values equal to 4.94. The second estimator returns random numbers from uniform distribution on [1.00, 9.00]. Expected mean AEs for valence and arousal, after rounding, are 2.27 and 1.47 for the first estimator and 2.76 and 2.32 for the second, based on the following derived formulas:

$$\mathrm{E}_{\text{unbiased constant}}[\text{mean AE}] = \frac{1}{N}\sum_{i=1}^{N} |X_i - \overline{X}|\,, \tag{1}$$

$$\mathrm{E}_{\text{uniform random}}[\text{mean AE}] = \frac{1}{8N}\sum_{i=1}^{N} (X_i^{\,2} - 10X_i + 41)\,. \tag{2}$$

In the formulas, N stands for the number of estimated valence/arousal outputs, X_i represents valence, or arousal, in the corresponding stimuli annotations, and $\overline{X}$ is the mean value of all X_i. Therefore, valence mean AE from Table 2 is about two times lower, and arousal mean AE is about two or three times lower, than the corresponding expected mean AEs of the two simplified estimators.

7 Conclusion

The paper has presented design and evaluation of the real-time Emotional State Estimator, using both artificial subjects and real humans. Inconveniences of using human subjects in early development and testing of the physiology-driven adaptive VR stimulation have been among the motivating factors for development of artificial subjects. Estimation accuracy is superior for artificial subjects, which is expected and acceptable with regard to their purpose.

In future work, comparative analysis is planned between different emotional state estimation methods, like various modifications of ANNs, hidden Markov models etc., different feature sets, as well as additional sensor inputs, like EEG, EMG etc. Larger homogenous samples of human subjects and more sophisticated fusion of multiple estimators' outputs might also improve real-time emotional state estimation. However, in order to emphasize human individuality, some form of estimator customization to each individual will also be applied.

Acknowledgements. This research has been partially supported by the Ministry of Science, Education and Sports of the Republic of Croatia. We thank Dr. Ivica Kopriva for valuable comments and Dr. Velimir Išgum for assistance with physiological measurements.

References

1. Ćosić, K., Popović, S., Jovanovic, T., Kukolja, D., Slamić, M.: Physiology-Driven Adaptive VR System: Technology and Rationale for PTSD Treatment. Annual Review of CyberTherapy and Telemedicine 5, 179–191 (2007)
2. Popovic, S., Slamic, M., Cosic, K.: Scenario Self-Adaptation in Virtual Reality Exposure Therapy of Posttraumatic Stress Disorder. In: Roy, M.J. (ed.) Novel Approaches to the Diagnosis and Treatment of Posttraumatic Stress Disorder. NATO Security through Science Series - E: Human and Societal Dynamics, vol. 6, pp. 135–147. IOS Press, Amsterdam (2006)
3. Schuemie, M.J., van der Mast, C.A.P.G., Krijn, M., Emmelkamp, P.M.G.: Exploratory Design and Evaluation of a User Interface for Virtual Reality Exposure Therapy. In: Westwood, J.D., Hoffman, H.M., Robb, R.A., Stredney, D. (eds.) Medicine Meets Virtual Reality 02/10, pp. 468–474. IOS Press, Amsterdam (2002)
4. Gerardi, M., Rothbaum, B.O., Ressler, K., Heekin, M., Rizzo, A.: Virtual Reality Exposure Therapy Using a Virtual Iraq: Case Report. Journal of Traumatic Stress 21(2), 209–213 (2008)
5. Wood, D.P., Murphy, J.A., Center, K.B., Russ, C., McClay, R.N., Reeves, D., Pyne, J., Shilling, R., Hagan, J., Wiederhold, B.K.: Combat Related Post-Traumatic Stress Disorder: A Multiple Case Report using Virtual Reality Graded Exposure Therapy with Physiological Monitoring. Medicine Meets Virtual Reality 16, 556–561 (2008)
6. Russell, J.A.: A Circumplex Model of Affect. Journal of Personality and Social Psychology 39, 1161–1178 (1980)
7. Lang, P.J., Bradley, M.M., Cuthbert, B.N.: International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual. Technical Report A-6. University of Florida, Gainesville, FL (2005)
8. Bradley, M.M., Lang, P.J.: The International Affective Digitized Sounds, 2nd edn. (IADS-2): Affective Ratings of Sounds and Instruction Manual. Technical report B-3. University of Florida, Gainesville, FL (2007)
9. Horvat, M., Popović, S., Bogunović, N., Ćosić, K.: Tagging Multimedia Stimuli with Ontologies. In: MIPRO (submitted, 2009)

10. Popović, S., Horvat, M., Ćosić, K.: Generator of Audio-Visual Stimuli for Psychotherapy and Psychological Training. In: Kozarić-Kovačić, D. (ed.) 3rd Croatian Symposium on Stress-Related Disorders, p. 40. University Hospital Dubrava, Zagreb (2008) (in Croatian)
11. Popović, S., Đukić, Z., Križek, V., Kukolja, D., Ćosić, K., Slamić, M.: Control of Emotional Characteristics of Stimuli in Virtual Reality Based Psychotherapy. In: Budin, L., Ribarić, S. (eds.) Proceedings of the 30th Jubilee International Convention MIPRO. CTS & CIS, vol. 3, pp. 91–96. MIPRO, Rijeka (2007)
12. Bradley, M.M., Lang, P.J.: Emotion and Motivation. In: Cacioppo, J.T., Tassinary, L.G., Bernston, G.G. (eds.) Handbook of Psychophysiology, 3rd edn., pp. 581–607. Cambridge University Press, Cambridge (2007)
13. Dawson, M.E., Schell, A.M., Filion, D.L.: The Electrodermal System. In: Cacioppo, J.T., Tassinary, L.G., Bernston, G.G. (eds.) Handbook of Psychophysiology, 3rd edn., pp. 159–181. Cambridge University Press, Cambridge (2007)
14. Lim, C.L., Rennie, C., Barry, R.J., Bahramali, H., Lazzaro, I., Manor, B., Gordon, E.: Decomposing Skin Conductance into Tonic and Phasic Components. International Journal of Psychophysiology 25, 97–109 (1997)
15. ECGSYN - A realistic ECG waveform generator, `http://www.physionet.org/physiotools/ecgsyn/`
16. McSharry, P.E., Clifford, G.D., Tarassenko, L., Smith, L.A.: A Dynamical Model for Generating Synthetic Electrocardiogram Signals. IEEE Trans. Biomed. Eng. 50(3), 289–294 (2003)
17. Lisetti, C., Nasoz, F.: Using Noninvasive Wearable Computers to Recognize Human Emotions from Physiological Signals. EURASIP Journal on Applied Signal Processing 11, 1672–1687 (2004)
18. Wagner, J., Kim, J., André, E.: From Physiological Signals to Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification. In: IEEE International Conference on Multimedia and Expo, pp. 940–943 (2005)
19. Rani, P., Liu, C., Sarkar, N., Vanman, E.: An Empirical Study of Machine Learning Techniques for Affect Recognition in Human-Robot Interaction. Pattern Analysis & Applications 9, 58–69 (2006)
20. Healey, J.: Wearable and Automotive Systems for Affect Recognition from Physiology. Technical report, Department of Computer Science, MIT, Cambridge, MA (2000)
21. Kim, K.H., Bang, S.W., Kim, S.R.: Emotion Recognition System using Short-Term Monitoring of Physiological Signals. Medical & Biological Engineering & Computing 42(3), 419–427 (2004)
22. Haag, A., Goronzy, S., Schaich, P., Williams, J.: Emotion Recognition using Bio-Sensors: First Step Towards an Automatic System. In: Affective Dialogue Systems, Tutorial and Research Workshop, Kloster Irsee, Germany, June 14-16 (2004)
23. Leon, E., Clarke, G., Callaghan, V., Sepulveda, F.: A User-Independent Real-Time Emotion Recognition System for Software Agents in Domestic Environments. Engineering Applications of Artificial Intelligence 20(3), 337–345 (2007)
24. Kaya, N., Epps, H.H.: Relationship between Color and Emotion: A Study of College Students. College Student Journal 38(3), 396–405 (2004)

User's Motion for Shape Perception Using CyARM

Ryo Mizuno, Kiyohide Ito, Tetsuo Ono, Junichi Akita, Takanori Komatsu, and Makoto Okamoto

[1] Graduate School of Systems Information Science, Future University-Hakodate, Japan
[2] School of Systems Information Science, Future University-Hakodate, Japan
[3] Kanazawa University Graduate School of Natural Science and Technology Division of Electrical Engineering and Computer Science, Japan
[4] Fiber-Nanotech Young Researcher Empowerment Center, Shinshu University, Japan
{g3107005,itokiyo,tono}@fun.ac.jp, akita@is.t.kanazawa-u.ac.jp, tkomat@shinshu-u.ac.jp, maq@fun.ac.jp

Abstract. We have developed a new sensing device, named by "CyARM". CyARM is one of Active Perception device, based on activeness of human perception. In order to clarify information for shape perception using CyARM, an experiment had conducted. Results of this experiment show that shapes with sharp edges are identified easily, but shapes with smooth edges are identified with greater difficulty. In other words, the subjects perceived changes in distance from the sensor to the object as object edges. Furthermore, from the results of multi-dimensional scaling, it is suggested that object shapes perceived by the subjects were classified according to the sharpness of the edge and by the ratio of height and width. In addition, motion analyses were conducted. The result shows, it is suggested that user tend to swing arm to front arm mainly in motion of lateral direction.

Keywords: Active Perception Device, CyARM, and Shape Perception.

1 Introduction

In recent years, with developments in sensor technologies sensing devices for assisting people to perceive their environments have developed. Such devices get data and provide it to the user. Sensing devices need to provide this data as meaningful information to users, because user cannot perceive environment from only such law data. Therefore, sensing devices have to be designed in view of human perceptions.

In this article, it is investigated about next two points through introduction of sensing device, which is thought human perception, and shape perception experiment using sensing device. 1) Information for shape perception using a sensing device. 2) How to move a sensing device for shape perception using a sensing device.

2 Concept

In this article, we focus on active human perception. Sensing devices for assisting people to perceive their environments are used at their daily environments. Such

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 185–191, 2009.

environment has complexity and change sequentially. Sensing devices need to provide this complex data as meaningful information to users. So that, in our study, ecological approach to human perception is adopted. In this approach, complex environments are intended.

2.1 Activeness of Perception

In the ecological approach, human perception is known as ‘active’ [1]. This means that the action is comprehended in the perception [2]. For example, head moving to see or listen surroundings is action for perception. For another example, in order to detect and perceive an environment, humans stretch arms, make sounds and listen to reflected sounds (echolocation). In other words, human pick up the information within the environment employing sensory organs (eyes, ears, and arms, et al.) for various purposes [3]. Therefore sensing devices, which help user’s perceptions, should be movable in ways that mimic sensory organs.

In this approach, it is thought the advantage of moving sensory organ is that such action can change the energy structure in the environment. By changing the energy structure, human can pick up the invariant. The invariant is information for perceive the environment.

2.2 Active Perception Device

We thought sensing devices, which help user’s perceptions, should be movable in ways that mimic sensory organs under activeness of human perception. Users may be able to pick up invariant to change sensing data using such sensing devices, and perceive their environment.

We called such sensing device, which user can move sensors freely, “Active Perception Device”[4].

2.3 CyARM

Over the past few years our research team has been developing a new sensing device, which we refer to as CyARM [5], to be applied specifically for the environment perception of visually impaired persons.

CyARM measures the distance between a user and an object by ultra sonic sensor and transmits it to the wire’s length, which connects to body of CyARM and user’s waist. Accordingly, distance data is offered as bending user’s arm. CyARM is Active Perception Device, which is able to be used changing sensing data.

2.4 Shape Perception Using CyARM

CyARM does not only measures and transmits the distance between a user and an object to the user’s haptic sense, but also helps perceive the shape of objects without viewing of the object (See Fig. 2). CyARM also extends arm movements so that the user can shake or walk with the device to perceive the shapes of the object. With the successive measurement of distance, CyARM records different distances when the user changes position or direction. It is assumed, hence, that the pattern in changing measured distances depends on the shape of object. From these points we

Fig. 1. CyARM

hypothesized that CyARM may enable the user to identify the shape by moving the device to trace the surface.

If user can perceive object shape using CyARM, what is the information for shape perception? It is thought that changing pattern of distance from sensor to object is able to be shape specific information. Therefore, user may be able to perceive object shape to pick up shape-specific changing pattern of distance from sensor to object.

The changing pattern from sensor to object is depends on two things following below.

1. Object shape
2. Way to move CyARM by user

In object shape, especially, object edges make changing of distance from sensor to object. There are two kinds of edges: sharp edge and rounded edge. At sharp edge, distance may change immediately. By contrast, at rounded edge, distance may change slowly. Therefore, edge kind is expected to be cue for shape perception using CyARM.

Fig. 2. Concept of Shape Perception using CyARM

3 Purpose

We conducted an experiment to prove the perceptibility of CyARM, to clarify information in shape perception, and to clarify the relationship between the way of moving

CyARM and the accuracy of shape perception by using it. Changes of distance by object shapes were focused on, specially, features of perceived shapes and classifying them.

4 Method

4.1 Subject and Material

Subjects were eight university students and two graduate students. They have never used CyARM by this experiment. Subjects ware blindfolded and equipped with mounted headphone.

As a device to perceive object's shapes, CyARM was adopted. In order to maintain distance from a subject and the object, the "guide" for locomotion was put into the place. The guide was 1.5m in radius. There were four shapes: "sphere", "cube", "rectangular solid", and "cylinder". Objects surfaces were common material (papers), because ultrasonic reflectance depends on object's surface material. All objects volume was 47713 cm^3. This experiment condition can be seen in Fig. 3. Furthermore, in this experiment, user's motions were captured by two video cameras (30 frame/s) and motion capture system (PV Studio 3D).

Fig. 3. Experiment Environment

4.2 Procedure

1. Subjects were asked to use CyARM and scan an object in a distant place. They were allowed to move only within the guide.
2. Then the subjects were required to state their judgments of the shape. There were four shapes: "sphere", "cube", "rectangular solid", and "cylinder". The objects were presented randomly.
3. The 12 trials consisted of three trials at each of the four objects. Before trials, there were training trials. In training trials, there ware two shapes: "cone" and "triangle pole". Subjects were informed start and end of trials by knocked their shoulders.

5 Result Specifying Shapes

The mean percentage correct identification score for the sphere, cube, rectangular solid, and cylinder were 26.7, 46.7, 40.0, and 20.0, respectively. A percentage response matrix for this experiment can be seen in Table 1. T tests compared the mean percentage of correct identification for each shape relative to the chance level (25.0%). Results of the T test show that performance with cube and rectangular solid was better than chance level at $p < .05$. By contrast, performance with sphere and cylinder was not significantly better than chance.

Table 1. Percentage of responses for each shape

Reported	Presented shape			
shape	sphere	cube	rectangular	cylinder
sphere	26.70	20.00	36.70	16.70
cube	10.00	46.70	23.00	20.00
rectangular	17.00	20.00	40.00	23.30
cylinder	40.00	30.00	10.00	20.00

5.1 Classification of Perceived Shapes

An additional analysis was conducted to determine similarities of identified shapes by multi-dimensional scaling. Result of this analysis, cube and rectangular solid are laid out in the superior region, and sphere and cylinder are laid out in the inferior region along vertical axis (dimension 1). Cube and sphere are laid out in the right side, and rectangular solid and cylinder are laid out in the left side along vertical axis (dimension 2). This chart can be seen in Fig. 4.

5.2 User’s Motion

Motion analysis was conducted to clarify the relationship between the way of moving CyARM, and the accuracy of shape perception by using it. The motion data of each arm’s joints (wrist, elbow, shoulder) and a sensor of CyARM while shape perception tasks were analyzed.

In order to analyze motion relationship between each arm’s joint (sensor-wrist, wrist-elbow, elbow-shoulder), each cross-correlation coefficient between time series of them were calculated. The mean of square values of cross correlation coefficient (r^2) are used to ANOVA for each pairs of arm’s joints and a sensor. the cross correlation coefficient of elbow-shoulder pair (0.70) was larger than sensor-wrist pair (0.61) and wrist-elbow pair (0.59). An ANOVA revealed significant main effects for each arm’s joints and sensor, $F(3, 10) = 3.83$, $p < .05$. Pairwise comparison tests for each pairs of joint showed that the cross correlation coefficient of elbow-shoulder pair was larger than wrist pairs significantly ($p< .05$).

The mean of square values of cross correlation coefficient in correct identification for the sensor-wrist pair, wrist-elbow pair, and elbow-shoulder were 0.69, 0.66, and 0.68, respectively. On the other hand, the mean of square values of cross correlation coefficient in erroneous identification for the sensor-wrist pair, wrist-elbow pair, and elbow-shoulder were 0.58, 0.57, and 0.70, respectively.

6 Discussion

6.1 Accuracy of Specifying Shapes

Performance results for the cube and rectangular solid were higher than random chance. In contrast, performance results for the sphere and cylinder were not significantly higher than random chance. These results show that shapes with sharp edges (cube and rectangular solid) are identified easily, but shapes with smooth edges (sphere and cylinder) are identified with greater difficulty. In other words, the subjects perceived changes in distance from the sensor to the object as object edges. Gradual changes in distance were not easily specified.

The low accuracy of the "cube" and "rectangular solid" testing may have been influenced by the fact that the subjects of the experiment had no prior experience using CyARM. Further study is needed to determine whether improvement in accuracy can be achieved after becoming acquainted with the use of CyARM.

6.2 Classification of Perceived Shapes

From the results of multi-dimensional scaling, it is suggested that object shapes perceived by the subjects were classified according to the sharpness of the edge and by the ratio of height and width. It is also possible to classify actual shapes in this manner. Therefore, these results may indicate no qualitative difference between identified shapes and actual shapes.

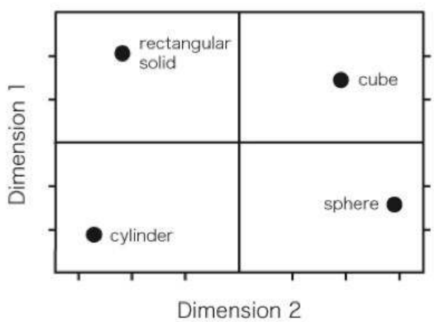


Fig. 4. Similarity of Shape identified by participants

6.3 User's Motion

If user swing elbow, the relationship between his elbow and his shoulder is not constant, and elbow-shoulder pair's cross-correlation coefficient is lower value. From the result of motion analysis, elbow-shoulder pair's cross-correlation coefficient was higher value than other pairs of body parts. Therefore, it is suggested that user tend to swing arm to front arm mainly in motion of lateral direction.

7 Summary and Future Works

In this article, we attempted to test whether a shape could be perceived using an "Active Perception Device" that moves similar to a sensory organ. In the experiment, the active perception device called CyARM was used. Results of the experiment showed that sharp-edged shapes could be identified, but smooth-edged shapes were more difficult to identify. Object shapes were classified by sharpness of the shape's edge and by the ratio of height and width.

As previously mentioned, identification of the shape of an object is determined not only by the change in distance between the user and the object, but also by how the user moves CyARM. Perceived shapes do not depend on only how to move, because percentage correct identification scores was not significantly difference. As such, continued study of the CyARM device is needed to analyze the relationship between the method of moving CyARM and the accuracy of shape perception, as well as the perception of shape itself, using the motion data between shape perception tasks.

Furthermore, motion analysis was conducted to clarify the relationship between the way of moving CyARM, and the accuracy of shape perception by using the motion data of each arm's joints and sensor of CyARM between shape perception tasks. In order to analyze motion relationship between each joint, each cross-correlation function between time series of them were calculated. As a result, he maximum value squared of cross-correlation coefficient between wrist and elbow motion of horizontal axis were marked difference. This result suggests that user swing arm to front arm mainly in motion of lateral direction.

We will conduct additional experiment in order to clarify effective motion for shape perception using CyARM. In the experiment, it is compared user's motion between expert of shape perception using CyARM and novice's it.

These findings will be eventually applied to the designing of a more sophisticated environment perception device. CyARM also is able to use as sensory substitution system for blind people. The information for shape perception using CyARM and effective way to move sensor of it for blind people and for general people would be not the same. In order to apply CyARM for sensory substitution system, we will.

References

1. Gibson, J.J.: Observation on Active Touch. Psychological Review 69(6), 491 (1962)
2. Gibson, J.J.: The Senses Considered as Perceptual Systems. Houghton Mifflin, Boston (1966)
3. Reed, E., Jones, R.: Reasons for Realism. Lawrence Erlbaum Associates, Mahwah (1982)
4. Mizuno, R., Ito, K., Akita, J., Ono, T., Komatsu, T., Okamoto, M.: Shape Perception using CyARM - Active Sensing Device. In: Proceedings of the 6th International Conference of Cognitive Science (ICCS 2008), pp. 182–185 (2008)
5. Okamoto, M., Akita, J., Ito, K., Ono, T., Takagi, T.: CyARM: Interactive Device for Environment Recognition using a Non-Visual Modality. In: Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D. (eds.) ICCHP 2004. LNCS, vol. 3118, pp. 462–467. Springer, Heidelberg (2004)

Human Control Modeling Based on Multimodal Sensory Feedback Information

Edwardo Murakami[1,⋆] and Toshihiro Matsui[2]

[1] Digital Human Research Center
National Institute of Advanced Industrial Science and Technology(AIST),
2-41-6, Aomi, Koto-ku, Tokyo 135-0064, Japan
edwardo.murakami@aist.go.jp
[2] Research Coordinator (Information Technology and Electronics)
National Institute of Advanced Industrial Science and Technology(AIST),
1-1-1 Umezono Tsukuba, Ibaraki 305-8568 Japan

Abstract. In order to simulate the human control behavior during a manipulation task in a remote controlled or in a X-by-wire systems, first it is necessary to measure and analyze the human control characteristics. The aim of this research is to measure the operator reaction time and analyze the human visual and force sensory feedback integration related to a manipulation task. Using the developed master-slave type experimental device it was possible to identify and build a human operator control model related to different sensory feedback. The human model related to visual feedback solely and visual/force feedback was identified using the techniques of system identification methods.

Keywords: Human-Machine Interface, System Identification, Reaction Time, Sensory Feedback Information.

1 Introduction

In order to simulate the human control behavior during a manipulation task in a remote controlled or in a X-by-wire systems, first it is necessary to measure and analyze the human control characteristics. The aim of this research is to analyze the human control characteristics in respect to the visual, force and audio feedback information and build a human control model that can also represent a control strategy based on multi-sensory feedback. This control model would be useful to assist the design, simulation and evaluation of human-machine systems like telerobots [1] and also computer assisted systems as power-assist and drive-by-wire vehicles. In this work the measurement of the reaction time will be compared to the time delay identified during a visual tracking task using only visual feedback and with force feedback information. An experiment device capable of measuring the human control characteristics in the presence of different sensory feedback information was developed.

⋆ Corresponding author.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 192–201, 2009.

2 Human Control Model

Many studies in the area of visual tracking have been done to understand the human control characteristics due to visual feedback information. However, quite a few researches have been conducted in the field of force and audio feedback information with the objective of analyzing how the human operator makes use of this sensory feedback information in a manipulation task.

2.1 Visual Feedback

One result of McRuer [2] [3] works about the analytical theory on manual control of vehicles was the Crossover Model. See Eq. 1 and Fig. 1. According to the manipulated machine characteristics the human operator can modify his/her own dynamic characteristics so as the open-loop transfer function remains a first order system.

$G(s)$ represents the machine dynamics.

$$H(s)G(s) \approx \frac{\omega_c e^{-\tau s}}{s} \qquad (near \;\; \omega_c) \tag{1}$$

where, τ ($0.1 \sim 0.4$s) represents the time lag due to human responses, ω_c ($0.5 \sim 0.8$Hz) is the crossover frequency.

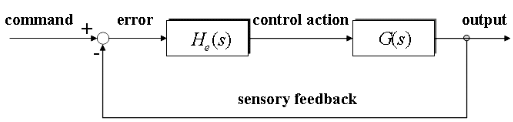


Fig. 1. Human Machine Block System

2.2 Force Feedback

The force feedback felt by the human operator is a result of a combination of tactile sensors and proprioceptive feedback. Although the individuals properties of each sensor have been studied, how the human operator uses those information and how they affect the human control characteristics are still not well known. However, it is of general agreement that the force feedback information is very important to identify the controlled object dynamics properties.

2.3 Audio Feedback

A primary function of audio feedback is said to direct the eyes to the source of the sound. More specifically in a tracking task the audio feedback provides information about the localization and velocity of the moving target. Although the space discrimination of auditory localization is not so accurate (about 15 degrees) compared to the visual, it provides supplementary information to assist other sensory feedback.

2.4 Human Control Model Based on Multimodal Sensory Feedback

This work proposes a human control model based on multiple sensory feedback information. (See Fig. 2). The human control characteristics related to visual, force and audio feedback information will be measured separately and then a combination of different sensory feedback will be analyzed in order to understand how the human operator uses these sensory feedback information to acquire an internal model of the controlled machine. First, in this work, the analysis of the human control strategy in the presence of visual, force and the combination of visual and force feedback information was conducted.

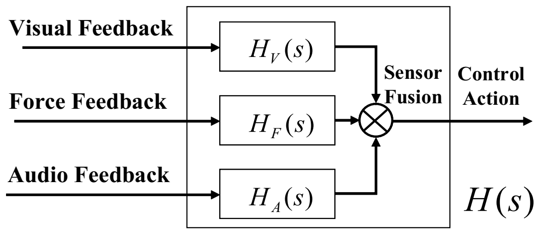


Fig. 2. Multimodal sensory feedback scheme

3 Experiments

Two types of experiments were conducted. First, the human reaction time was measured. The second experiment is based on visual tracking in the absence and presence of force feedback. The proposal for this study was reviewed by the Institutional Committee for Ergonomic Experiments and approved by the Director of Safety and Environmental Protection Department.

3.1 Reaction Time Experiment

Before conducting the operator model identification experiments, each subject's Single Reaction Time (SRT) and Choice Reaction Time (CRT) were measured to be compared to the time delay obtained from the second experiment. During this experiment the forearm pronator and supinator muscles EMG was also measured.

Single Reaction Time (SRT). In the SRT experiment the subject was instructed to turn the dial after the visual cue, which is a LED source, is presented. The instructed direction of rotation was clockwise (CW). The visual cue is presented in a random time between 3s and 9s after the go signal. After some training each subject executed 10 trials.

Fig. 3. SEesaw Experimental Device (SEED)

Choice Reaction Time (CRT). In the CRT experiment the subject has to turn the dial in the CW direction if the right visual cue is presented or in counter clockwise (CCW) direction if the left LED turns on. The side, right or left, and the visual cue presenting time is showed randomly to the subject. After become familiar with the task each subject performed 20 trials.

SRT and CRT Results. The results of the SRT and CRT experiments are shown in Table 1. In both SRT and CRT experiments the reaction time was defined as the time necessary to the subject rotate the dial more than 45 degrees after the visual cue was presented. The subject A was the fastest and the subject B had the slowest response. The difference between these 2 subjects can be attributed to the time necessary to perceive the visual cue and send the motor commands to the muscles since the starting of EMG activation differs greatly between subjects. (See Fig. 4 and Fig. 5). The time necessary to process the visual cue information and send motor command to the muscles varied from 0.15s to 0.29s in SRT experiment and was between 0.19s to 0.34s in CRT experiments. Thus it can be inferred that the time need to decide which direction to move was between 0.04s to 0.15s. On other side, there were no much difference among subjects in the time between the muscle activation and movement onset which corresponds to 0.02s - 0.05s. These results demonstrate that the cognition and decision making are responsible for great part of the human response delay.

3.2 Operator Control Characteristics Identification Experiment

The method to model the human operator adopted in this study is based basically in the system identification used in control theory. However, it is crucial

Table 1. Results of SRT and CRT Experiments (mean ± SD)

Experiment	Subject A	Subject B	Subject C
SRT [s]	0.29 ± 0.03	0.38 ± 0.05	0.32 ± 0.03
CRT [s]	0.34 ± 0.04	0.46 ± 0.07	0.38 ± 0.05

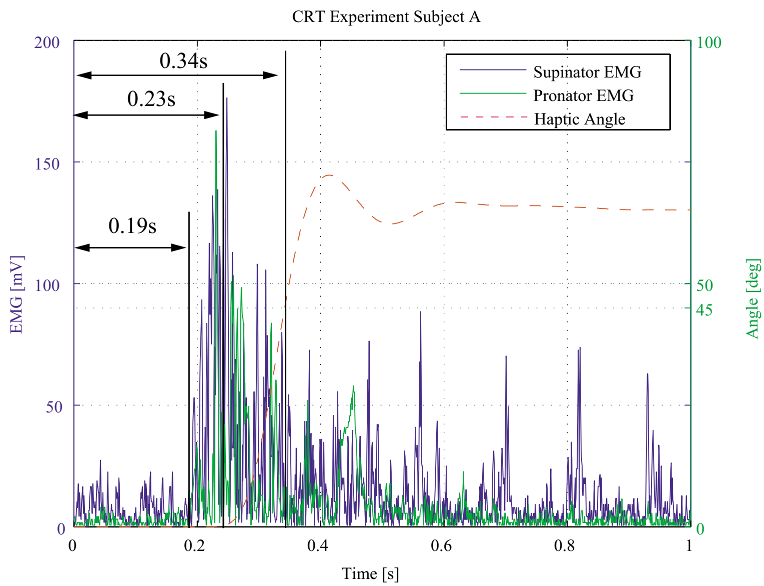


Fig. 4. CRT experiment subject A (fastest). The EMG activation before the movement actually started can be noticed.

to select a task that can provide an analysis of the operator characteristics in a visual and force control manipulation independently, i.e. a task that can be performed with only one type of sensory feedback information. It is also preferable to be a continuous task for system identification analysis in a wide frequency range. The peg-in-hole task is widely used as an example of robot control, but it is very hard to decompose the position and force control strategies. The inverted pendulum is also commonly used to demonstrate different control methods. However, it is a task very difficult to accomplish with the eyes closed. After considering many tasks performed by a human operator, the control of a slider on a seesaw was chosen as a suitable task that can pull together all the necessary features to analyze and identify the human-machine system related to visual, force and also audio feedback information independently. To analyze the human control characteristics related to different sensory feedback properties a master-slave type SEesaw Experimental Device (SEED) was developed. The master haptic device consists of a dial with a force sensor and the slave is an actuated linear guide that works as a seesaw bar with a slider over it. (See Fig. 3 and Fig. 6). After

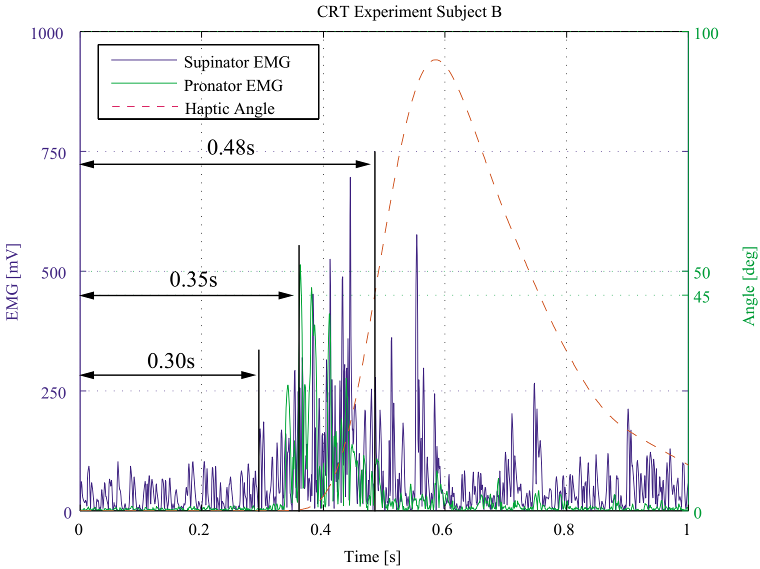


Fig. 5. CRT experiment subject B (slowest). The starting of the EMG activation ocurred later than the time compared to the fastest subject.

analyzing the control characteristics based on visual ($H_v(s)$), force ($H_f(s)$) sensory feedback information separately, the combination visual and force feedback ($H_{vf}(s)$) was analyzed.

Visual Pursuit Tracking with a Normal Slider. In this first experiment to analyze the human visual feedback properties, the human operator manipulates the master dial in order to make the slider, PD controlled, follow a random reference signal displayed in a monitor. The machine characteristics is a first order system. After 20 training trials, 10 trials were measured. Fig. 7 shows the reference signal, the measured data and the output of the identified operator model. The technique used to identify the human operator's characteristics is common to all the following two experiments. First, it was assumed that the human-machine open loop transfer function has the generalized form of Eq. (2).

$$H(s) = K\frac{(1+T_L s)}{(1+T_I s)}e^{-\tau s} \tag{2}$$

where K represents proportional gain, $e^{-\tau s}$: time delay due to human response, $(1+T_L s)$ is the lead time constant (relative rate-to-displacement), $(1+T_I s)^{-1}$ is the lag time constant.

Using the process model identification of Matlab toolbox the appropriate parameters were calculated by minimizing the error between the model output and the measured data. By this search the most suitable form was selected and then the time delay which corresponds to the smaller fitting error was explored. In all

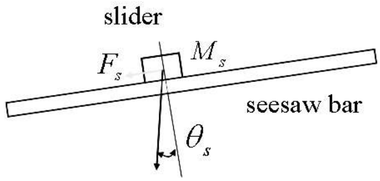


Fig. 6. Seesaw and slider model

the cases a priori knowledge about the controlled machine was used in order to obtain the operator characteristics. After obtaining the parameters, they were averaged separately and the results are shown in Table 2.

Visual Feedback using a Seesaw Task. Here the subject is instructed to follow the random reference signal as the previous experiment. But this time the machine dynamics behaves as a slider over a seesaw, i.e. like Eq. (3). There is no force feedback. After some practice the subject is able to execute successfully the task. To avoid the subject to notice that the slider behavior corresponds to a seesaw task, the seesaw bar was maintained in horizontal position.

$$(\theta_s \approx 0 \Rightarrow sin\theta_s \approx \theta_s) \qquad G(s) = \frac{x_s}{\theta_s} = \frac{g}{s^2} \tag{3}$$

where g is the gravity, θ_s is the seesaw bar inclination, x_s is the slider position and s is the Laplace operator

Visual and Force Feedback using a Seesaw Task. In this task the subject has to follow a random reference signal feeling the torque caused by the dislocation of the slider over the seesaw. The proficiency in the task execution was similar to the visual feedback solely, but the human modeling error increased due to the need of extra operational force. (Fig. 9).

3.3 Human Control Characteristics Experiment Results

The results of identified human control characteristics are shown in Table 2. The human control model related to visual pursuit task showed a first order

Table 2. Identified Human Control Model

Experiment	Subject A	Subject B	Subject C
Pursuit $H_p(s)$	$\frac{2.3}{1+0.4s}e^{-0.20s}$	$\frac{2.1}{1+0.6s}e^{-0.27s}$	$\frac{2.2}{1+0.5s}e^{-0.22s}$
Visual $H_v(s)$	$6\frac{1+4s}{1+0.06s}e^{-0.24s}$	$4\frac{1+4s}{1+0.2s}e^{-0.31s}$	$10\frac{1+2s}{1+0.03s}e^{-0.21s}$
Visual/Force $H_{vf}(s)$	$10\frac{1+2s}{1+0.04s}e^{-0.23s}$	$8\frac{1+2s}{1+0.1s}e^{-0.29s}$	$13\frac{1+2s}{1+0.04s}e^{-0.22s}$

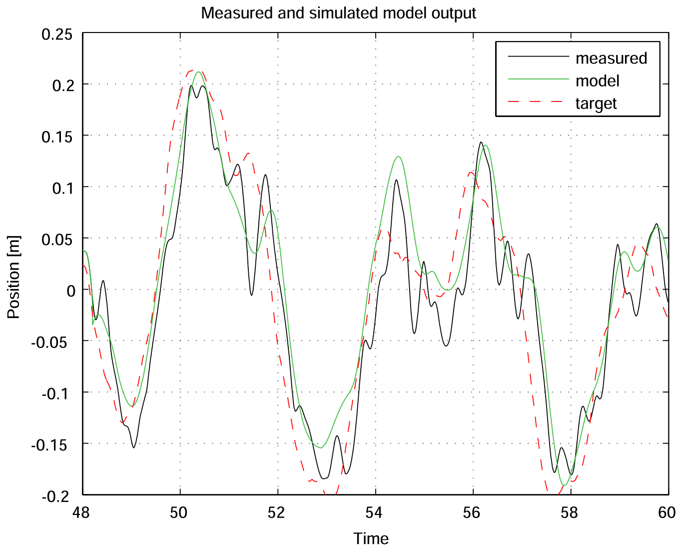


Fig. 7. Visual pursuit experiment

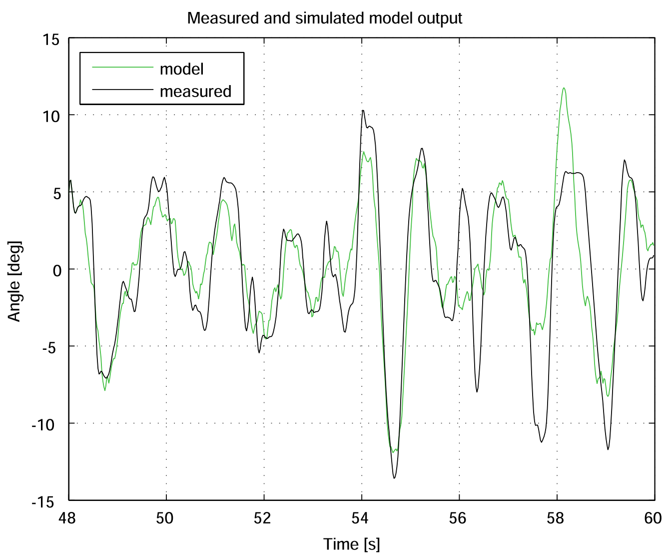


Fig. 8. Seesaw task experiment using only visual feedback information

characteristic. Comparing the three subjects it can be noticed that the lag time element T_I is proportional to the correspondent time delay τ . Fig. 7 shows the reference target, the measured data and the control model output of the identified mean human model. The model output has the same behavior of the human operator except the high frequency features.

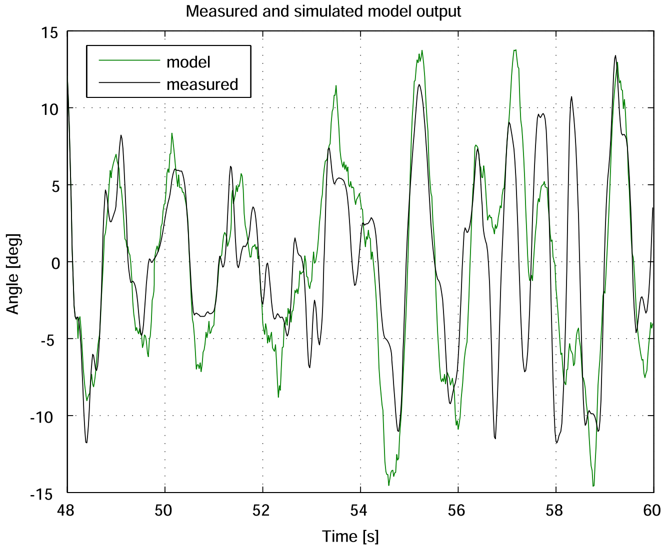


Fig. 9. Seesaw task experiment using visual and force feedback information

In the case of seesaw task using visual feedback with or without the presence of force feedback, all the subjects presented a lead time element T_L. This lead time element is responsible for the prediction of the slider's behavior. Due to the high acceleration of the gravity, a predictive element was necessary to make possible the control by the human operator.

The presence of force feedback had the effect of decreasing the time constant and increasing the gain element. The latter one can be attributed to the high stiffness of the forearm because of the haptic feedback.

4 Discussion

From Table 2, it can be noticed that according to the characteristics of the task the time delay identified has different values. But the relation between subjects is preserved as the subject B has the biggest time delay. Although the time delay is different from the RT experiment results, the direct measurement of the human response time presents a reasonable and practical method of identifying the time delay reducing the number of parameters to be fitted. Further investigation about the muscle activation time and neuromuscular dynamics should be conducted to achieve a better estimation of the human response delay.

5 Conclusion

This research proposed an analytical method using the SEED to identify the human control characteristics related different sensory feedback information. The human model related to visual feedback solely and visual/force feedback was identified using the techniques of identification methods. It is important to notice that all the experiments were performed without audio information. The next step is to build a human sensory feedback integration model to represent the human operator including also the audio feedback information. These sensorial feedback information are believed to play an important role in the acquisition of the internal model of manipulated machines. Future work will be done to analyze how the human model could be decoupled in feedforward, representing the internal model, and feedback elements.

References

1. Sheridan, T.B.: Telerobotics, Automation, and Human Supervisory Control. MIT Press, Cambridge (1992)
2. McRuer, D.T., Jex, H.R.: A Review of Quasi-Linear Pilot Models. IEEE Trans. Human Factors in Electronics HFE-8(3), 208–231 (1967)
3. McRuer, D.T.: Pilot-Induced Oscillations and Human Dynamic Behavior, NASA Contractor Report 4683 (1995)
4. Schenk, T., Mai, N.: Time constraints improve reaching movements in an ataxic patient. Experimental Brain Research 128(1-2), 214–218 (1999)

Potential and Challenges of Body Area Networks for Affective Human Computer Interaction

Julien Penders, Bernard Grundlehner, Ruud Vullers, and Bert Gyselinckx

Holst Centre / IMEC-NL, High Tech Campus 31, 5656 AE Eindhoven, The Netherlands
{Julien.Penders,Bernard.Grundlehner,Ruud.Vullers,
Bert.Gyselinckx}@imec-nl.nl

Abstract. The Human++ program aims at achieving highly miniaturized, wireless, intelligent and autonomous body sensor nodes to assist our health, comfort and lifestyle. In this paper the concept of body area network is applied to wireless monitoring of emotions, thus opening a new, affective, dimension in human computer interaction. A prototype body area network targeting the monitoring of physiological responses from the autonomous system is introduced, and tested for the classification of discrete emotions. Using data fusion and regression analysis, we show that the wireless physiological data can be mapped in real-time to an estimation of an individual's arousal level. Results in a controlled environment are presented, and specific challenges that need to be overcome for a widespread use of the technology are discussed. Finally, we show how advances in micro-power generation devices may lead to fully autonomous systems in the future.

Keywords: Ambulatory, Body area networks, Emotion monitoring, Ultra-low-power, Wireless.

1 Introduction

It is anticipated that micro and nano-system technology will increase the functionality of lifestyle and healthcare devices to gradually match the needs of society. It is expected that, in the next decade, technology will enable people to carry their personal body area network (BAN) that provides medical, lifestyle, assisted living, sports, entertainment and computer interface functions for the user. This network comprises a series of miniature sensor/actuator nodes, implanted or located at the body surface. Each node has its own energy supply, consisting of storage and energy scavenging devices. Each node has enough intelligence to carry out its task. Furthermore, each node is able to communicate with other sensor nodes or with a gateway node worn on the body. The gateway node communicates with the outside world using a standard telecommunication infrastructure such as a wireless local area or cellular phone network. On the other extremity of the network, experts then provide services to the individual wearing the BAN. Intelligent or expert systems further include data fusion algorithms for the aggregation of body sensor data into metrics quantifying an individual's health status, his physical, cognitive and emotional state. Next generation of

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 202–211, 2009.

BAN will include feedback loops for health, performance or stress management, and may enable a new, affective, dimension in human computer interface.

Early deployment of technology in different application cases are translated into critical technology obstacles that will need to be solved. The Human++ research program tackles key technology challenges associated to micro-power generation and storage, ultra-low-power radios, ultra-low-power DSPs, sensors and actuators [1]. The ultimate target is the development of miniaturized body sensor nodes, truly non-invasive, capable of data analysis and wireless communication and powered by body-energy.

In this paper the use of body area network technologies to enable affective human computer interface is discussed. In the following sections we will present a body area network for ambulatory monitoring of physiological responses from the autonomous nervous system, and show how this platform can be used to monitor an individual's arousal level in real-time. Furthermore, we will highlight key challenges that need to be addressed in the coming future. Finally, we will show how recent advances in micro-power harvesters enable autonomous wireless physiological monitors. A world which adapts to one's emotions and feelings may not be that far out anymore…

2 Enabling Ambulatory Wireless Emotion Monitoring

Monitoring of emotions or mental health has received a special interest in the last few years. In this context, emotion is usually defined as a mental and physiological state associated with a wide variety of feelings, thoughts, and behaviors. Following this definition, monitoring physiological and cognitive signals should enable to understand and "read" the emotional state of an individual in a particular situation. When based on non-intrusive measurements, this will enable new ways of human-machine interaction, and therewith a new range of applications in the domains of (mental) health management, safety, entertainment and ambient intelligence.

Emotion can be defined in terms of discrete emotional states [2] or as a position on the 2-dimensional arousal-valence space [3], or as a combination of both [4]. A number of groups have reported a wide range of studies to the objective evaluation of emotions, investigating varying modalities such as facial expressions [5], vocal patterns [6, 7], physiological responses [8, 9] or combinations of the above [10]. As a starting point to better capture technology opportunities and challenges for body area networks when applied to mental health monitoring and affective HCI, we have chosen to focus on monitoring physiological responses from the Autonomic Nervous System.

2.1 Body Area Network for Monitoring Autonomic Nervous System Responses

We recently reported the realization of a low-power body area network for monitoring ECG, respiration, skin conductance and skin temperature [13]. Each of these modalities is known to be regulated by the Autonomic Nervous System, and thus represent interesting candidates to capture ANS responses to external stimuli. The system, illustrated on Fig. 1, consists of two low-power miniaturized body sensor nodes which communicate with a receiver connected to a pc or to a data logger. The first node is integrated in a wireless chest belt and monitors ECG (lead-I) and respiration. The second node is integrated in a wireless wrist sensor and monitors skin conductance and skin temperature. The total size of each individual node is approximately 40 x 25 x 8 mm^3, including battery, sensors and read-outs. Power consumption of the ECG/respiration node is 2.5mA, whereas the wrist-based sensor consumes 4mA, mainly due to an infra-red temperature sensor. Low-power and high performance ECG monitoring is achieved through the use of a proprietary single channel ASIC for biopotential read-out [11]. The ASIC consists of AC coupled chopped instrumentation amplifier, a spike filter, and amplification stage with constant gain, and a variable gain amplifier stage. The variable gain amplifier can be used to electronically adjust the gain of the readout for varying needs of EEG, ECG and EMG applications. Power consumption of the ASIC is 60uW, leading to an average consumption of 75 uW for the ECG read-out.

Fig. 1. Integrated body area network for ambulatory monitoring of physiological responses from the Autonomic Nervous System

This body area network for ANS responses monitoring has been tested in controlled environment to evaluate its potential usage for monitoring emotional states [13]. 10 subjects (mean age 29.3, 3 females, 7 males) are involved in the study. They are asked to watch 5 emotionally arousing film clips to elicit sadness, happiness, fear, disgust and neutrality [12], while wearing the wireless monitoring equipment. At the end of each clip, the subjects fill a self-report questionnaire. The 5 film clips are grouped into 3 categories in function of their expected arousal level: fear and disgust, happiness and neutral, and sadness. The four physiological signals are analyzed offline, and a set of 13 features are extracted based on general physiology considerations and previous studies on emotion recognition [16]. The 13 features are then mapped to 2 axes using Fisher Mapping. Linear Discriminant Classification is finally used to classify the data. This process eventually leads to error rates of 0.36 computed using leave-one-out cross-validation on the data-set, which is similar to previous studies on emotion classification [9, 13, 15].

2.2 Real-Time Arousal Monitoring

In a second study, we have investigated the possibility to use the proposed BAN system as an enabling technology platform for performing real-time measurement of an individual's arousal level [17]. 20 healthy volunteers are involved in the experiment. A movie extract is chosen as the arousal stimulus, characterized by a calm beginning followed by a building-up phase culminating to a frightening event. A reference or target arousal function is defined as being zero during most of the movie, except in a region surrounding the frightening event (see [17] for all details on the choice of the target function). Volunteers are asked to watch the movie while their physiological signals are monitored using the wireless system. All tests have been performed in a controlled laboratory environment, in order to minimize the sources of distraction that may eventually lead to unexpected and uncontrolled increases in arousal. A set of features is extracted from the ECG and skin conductance signals, found to be the most responsive parameters to the tests. In a second step, these features are combined in an optimal arousal estimator using linear regression against the target arousal level. The outcome of the regression analysis is a set of coefficients, characterizing the importance of each individual feature in the final estimation of arousal. All algorithms are implemented in the Matlab computing environment. All algorithms are real-time, such that the process of feature extraction and arousal estimation can be applied in real-time on incoming signals measured using the BAN system.

The resulting estimator has then been used to monitor the arousal level of individuals wearing the system. Several tests have been performed in various environments, varying from laboratory to public places. As much as possible, the test subject was isolated from the outside world, for instance by using headphones. The test protocol used for these tests consists of four parts: a short movie to get acclimatized, a modified Stroop test [], an audio extract and a movie fragment. The Stroop test is modified to induce confusion (and hence mental stress) in the second part of the test. The audio extract is a 3-minute very relaxing piece of classical music abruptly disturbed by noises of several kinds after 120 and 150 seconds, expected to trigger startling responses. The movie clip is identical to the one used to develop the arousal monitor. An example of the estimated arousal level over the test sequence is given in Fig. 2 or one of the test subjects. In this figure, the solid line gives the estimated arousal over time. The dashed, vertical lines represent the events, as specified by the name shown directly to the right of these lines. It can be seen from this picture that the subject did not show a significant increase in estimated arousal during the modified Stroop test. There was, however, a sharp and large increase in estimated arousal level just after the audio events and the movie event. Apart from the expected responses, there are also some responses that clearly do not origin in any of the events. These false positives can be due to anything that triggers the subject's mind, such as an arousal triggering thought, or something surprising in the surrounding environment.

Overall, the arousal monitor has proven to work quite reliability in controlled environment—that is, in a lab setting where distraction opportunities are minimized. It has also been shown that the conclusions can be generalized from movie to other arousing stimuli, as suggested in Fig. 2. However, further experiments are needed before conclusions can be drawn about the extension of the results to non-controlled

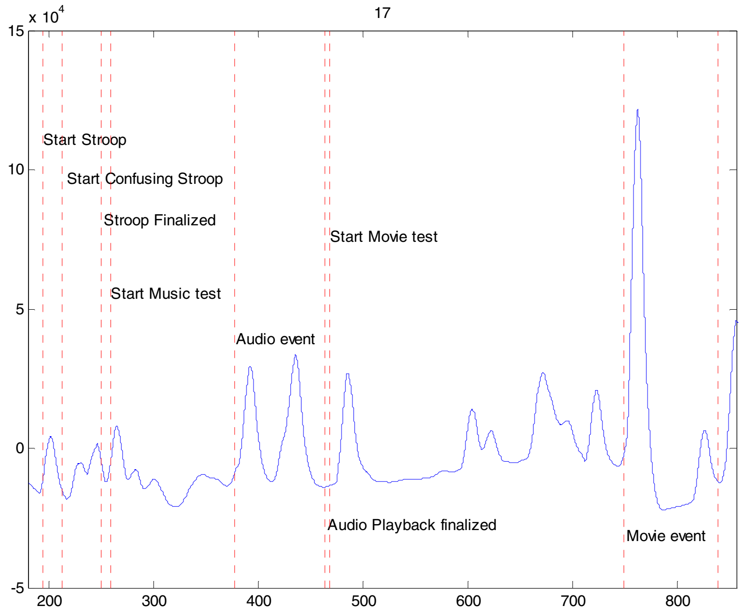


Fig. 2. Estimated arousal level over the part of the test protocol: modified Stroop test, audio extract and movie clip. Dashed vertical lines give the timing of the events.

environments. The low-power monitoring platform presented here certainly will facilitate the transition from lab to real-life environments.

2.3 Technology Challenges

The evaluation of the low-power body sensor network for ANS responses monitoring in different application cases offers new perspectives towards an objective, wireless and ambulatory monitoring of an individual's emotions. This technology evaluation exercise also leads to a better understanding of application requirements, and to the identification of important technology challenges that shall be addressed to eventually enable widespread deployment of body area networks for human computer interaction.

- **Ultra-low-power electronics:** current prototypes still mostly rely on low-power COTS components, from which the best lead to typical current consumption ranging from 1 to 10 mA, depending on the application. We previously showed that most of the power is consumed in the wireless transmission of the data, or in local processing of the data [1]. In some cases, sensors are also found to consume a significant part of the power. Further research is needed on ultra-low-power analog interfaces, sensors, DSP and radios. The ultimate target is to reach a total power consumption of 100 µW per body sensor node.
- **Autonomous systems:** the prototypes presented in this paper can run for a few days at full functionality. Breakthroughs in ultra-low-power technologies will

eventually enable months or years of autonomy. To come to a truly autonomous system however, it should be able to operate over its full lifetime without maintenance. Harvesting energy from the environment during the operation of the system will allow the system to run eternally with a battery or a supercapacitor acting only as a temporary energy buffer.

- **Multi-parameter sensors:** emotion monitoring is a complex problem, which is further reinforced by the fact that the underlying psycho-physiological aspects are not yet fully understood. Extending the range of functionality to include new sensing modalities will be crucial in fostering research in this area, leading to new discoveries. On the short term, multiple available sensors can be combined in the same system to enrich the information available, such as muscle tension, brain activity, voice, etc. On the longer term, novel sensing technologies are needed to reliably measure more complex parameters such as chemical compounds, hormones and proteins in body fluids, whilst pursuing ultra-low-power consumption. Continuous measurement of cortisol in saliva would for instance open new perspectives in stress monitoring.
- **Dry electrodes:** widespread acceptance of body area networks for human computer interface is expected to be intimately related to the comfort and easy of use of the system. Most of current systems for ECG, EMG and EEG monitoring require wet, gel electrodes to be attached to the skin or the scalp. Although they have the major advantage of providing good quality signals, gel electrodes exhibit significant drawbacks with regards to long-term use and ease of set-up. Dry electrode technology is required to enable simple setup of the system by the user itself. Several research groups have explored the area of dry electrode for EEG monitoring applications, for instance [19]. Further research is required to systemically tackle the issues of signal quality, robustness to motion artifact and bio-compatibility.ï
- **Increasing functionality:** most of today's body-worn sensors act as simple gateways, passing on the information to a central hub where the data is converted into actionable information. Emotion monitoring requires simultaneous monitoring of multiple sensors on the body, and the local extraction of relevant information out of the sensor data. Low-complexity and real-time algorithms are required to enable intelligent autonomous systems. Furthermore, compromises between local processing of the data versus data streaming or data storage exist and need to be investigated. A rational approach to distributed processing will allow achieving optimal performances for minimized power consumption.
- **Integration technology**: as body sensor nodes shrink in size and power consumption, end-user acceptance and compliance will eventually be bound to comfortable of use. Pioneering research in electronic integration technology has led to first functional prototypes of ultra thin chip packages [20] and stretchable interconnects [21]. Electronic integration in bi-dimensional flexible and stretchable foils will enable disappearing body sensor nodes, integrated in patches, clothes or even fashion accessories.

Addressing these technology challenges will lead to increased functionality, higher performance, better integration and decreased power consumption, thus bringing technology for ambulatory emotion monitoring closer to the end-user. In the next section we show how advances in micro-power generation systems already pave the way towards autonomous wireless sensor systems, enabling new perspectives for affective human computer interaction.

3 Towards Autonomous Wireless Health Monitors

The body is an under-estimated source of energy. It has been shown that the heat flow generated by the human body generates a power density of about 20 mW/cm^2 in average. This makes body thermal energy an interesting candidate for harvesting. Thermal-energy harvesters are thermoelectric generators which exploit the Seebeck effect to transform the heat flow from the human body to the environment into electrical energy, typically showing efficiencies of 0.1 to 3 %. Besides the energy dissipated by the body, much energy is also available from the surrounding environment, mainly from electromagnetic radiation (natural and artificial light). A recent survey of commercially available photovoltaic solar-cells has shown that indoor photovoltaic cells are capable of generating a maximum of 8 to 14 μW/cm^2 (depending on the type of cell) at perpendicular incidence and under 400 Lux. The type of light source was found to have only a minor effect on the generated power density.

In 2007, we reported the first autonomous health monitoring systems powered by thermal-energy harvesters: a wireless autonomous pulse-oximeter powered by a wrist-watch type TEG, and a wireless autonomous EEG monitoring system powered by a head-band type TEG [22, 23]. An important issue with these prototypes was the dependence of the generated power on the ambient temperature. To cope with this issue, we realized an improved prototype of autonomous wireless EEG monitor, featuring a hybrid power supply [24], as illustrated in Fig. 3. This power supply combines a TEG, which uses the heat dissipated from a person's temples, and Si photovoltaic cells. The TEG is composed of six thermoelectric units made up from miniature commercial BiTe thermopiles. Two high-efficiency Si photovoltaic cells are integrated on the left and right sides of the head, each of them having an area of 4x8 cm^2. These cells play the double function of converting ambient light into electricity, and serving as a part of the radiator to ensure effective heat transfer from the head into the environment. Exploiting the advantages of dual energy sources, the dimensions (size and weight) of the TEG have been reduced in comparison to previous prototypes, the power/volume ratio increased, and the range of ambient temperature at which the system works reliability widened.

To enable their use in electronic device, the TEG requires advanced power management circuitry to optimize harvested power efficiency. Typically, the TEG continuously charges a battery or a super-capacitor, which then provides power to electronic modules. Voltage up-converters are usually added to match the need for higher voltage power-supply of different electronic components. In parallel to the TEG power conversion circuit, a secondary power management circuit allows charging the battery directly from photovoltaic (PV) cells.

Fig. 3. Autonomous wireless EEG system powered by body heat and ambient light

The EEG system integrates a proprietary ultra-low-power biopotential readout ASIC [11], and the whole system consumes only 0.8 mW. The entire battery-free 2-channel EEG system is wearable and integrated into a device resembling headphones, as illustrated in Fig. 3.

This example shows that prototypes of autonomous health monitoring systems can be achieved today. Nevertheless, it also suggests that research in miniaturization of energy harvesters system using micro-machining techniques is necessary to further miniaturize the devices, and make them available at a reasonable cost. Furthermore, miniaturization is intrinsically related to power consumption, as the micro-power module, harvester and/or battery, is making up for most of the size of wireless sensor systems. Energy harvesting techniques for body sensor network can achieve a power density of 10 to 100 $\mu W/cm^2$, with today's TEG achieving 25 $\mu W/cm^2$ in average. The technology challenge will thus be to make the step from low-power electronics (10-100 mW) to ultra-low-power technologies (0.1-1 mW).

4 Conclusion

In this paper, the concept of body area network has been applied to wireless monitoring of physiological responses from the autonomic nervous system, paving the way towards wireless ambulatory monitoring of emotions. The proposed prototype integrates early technology achievements from the Human++ research program on wireless autonomous sensors. Using data fusion and regression analysis the wireless physiological data can be automatically analyzed in real time, allowing the determination of an estimated arousal level. The future will see the integration of additional sensor modalities, to enable real time monitoring of a person's emotional state as a combination of arousal and valence. Furthermore, widespread acceptance of body area networks for human computer interface will require overcoming key technology challenges in terms of ultra-low-power radios, DSPs and analog interfaces, dry electrode research and 2D flex/stretch electronic integration. In particular, this paper suggests how advances in micro-power generation devices may lead to fully autonomous systems in the future.

References

1. Penders, J., Gyselinckx, B., et al.: Human++: from technology to emerging health monitoring concepts. In: Proceedings of the 5th International Workshop on Wearable and Implantable Body Sensor Networks, June 1-3, pp. 94–98. The Chinese University of Hong Kong, China (2008)
2. Cacioppo, J.T., Berntson, G.G.: Relationships between attitudes and evaluative space: A critical review with emphasis on the separability of positive and negative substrates. Psychological Bulletin 115, 401–423 (1994)
3. Plutchik, R.: Emotion: Theory, research and experience, ch. 1. Theories of emotion, vol. 1, pp. 3–31. Academic Press, New York (1980)
4. Russell, J.A.: A circumplex model of affect. Journal of Personality and Social Psychology 39, 1281–1288 (1980)
5. De Silva, L.C., Miyasato, T., Nakatsu, R.: Facial emotion recognition using multimodal information. In: Proc. IEEE Int. Conf. on Information, Communications and Signal Processing (ICICS 1997), Singapore, September 1997, pp. 397–401 (1997)
6. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollilas, S., Fellenz, W., Taylor, J.: Emotion recognition in human-computer interaction. IEEE Signal Process. Mag. 18(1), 32–80 (2001)
7. Fernandez, R., Picard, R.W.: Classical and novel discriminant features for affect recognition from speech. In: Proc. of Interspeech 2005 (September 2005)
8. Kreibig, S.D., Wilhelm, F.H., Roth, W.T., Gross, J.J.: Cardiovascular, electrodermal, and respiratory response patterns to fear- and sadness-inducing films. Psychophysiology 44(5), 787–806 (2007)
9. Picard, R., Vyzas, E., Healey, J.: Toward machine emotional intelligence: analysis of affec-tive physiological state. IEEE Trans. Pattern Anal. Mach. Intell. 23(10), 1175–1191 (2001)
10. Bailenson, J.N., Pontikakis, E.D., Mauss, I.B., Gross, J.J., Jabon, M.E., Hutcherson, C.A., Nass, C., John, O.: Real-time classification of evoked emotions using facial feature tracking and physiological responses. International Journal of Human-Computer Studies 66, 303–317 (2008)
11. Yazicioglu, R.F., Merken, P., Puers, R., et al.: A 60 μW 60 nV/√Hz Readout Front-End for Port-able Biopotential Acquisition Systems. IEEE J. Solid-State Circuits, 1100–1110 (May 2007)
12. Gross, J.J., Levenson, R.W.: Emotion elicitation using films. Cognition and Emotion 9(1), 87–108 (1995)
13. Brown, L., Grundlehner, B., van de Molengraft, J., Penders, J., Gyselinckx, B.: Body Area Networks for monitoring autonomous nervous system responses (in press)
14. Lisetti, C.L., Nasoz, F.: Using noninvasive wearable computers to recognise human emotions from physiological signals. EURASIP Journal on Applied Signal Processing 11, 1672–1687 (2004)
15. Kim, K.H., Bang, S.W., Kim, S.R.: Emotion recognition system using short-term monitoring of physiological signals. Med. Biol. Eng. Comput. 42(3), 419–427 (2004)
16. Collet, C., Vernet-Maury, E., Delhomme, G., Dittmar, A.: Autonomic nervous system response patterns specificity to basic emotions. J. Auton. Nerv. Syst. 62(1-2), 45–57 (1997)

17. Grundlehner, B., Brown, L., Penders, J., Gyselinckx, B.: The design and analysis of real-time, continuous arousal monitor (in press)
18. Stroop, J.R.: Studies of interference in serial verbal reactions. Journal of Experimental Psychology 18, 643–663 (1935)
19. Ruffini, G., et al.: First human trials of a dry electrophysiology sensor using a carbon nanotube array interface. Sensors and Actuators A: Physical 144(2), 275–279 (2008)
20. Christiaens, W., Loeher, T., Pahl, B., Feil, M., Vandevelde, B., Vanfleteren, J.: Embedding and assembly of ultrathin chips in multilayer flex boards. J. Circuit World 34(3), 3–8 (2008)
21. Brosteaux, D., Axisa, F., Gonzalez, M., Vanfleteren, J.: Design and Fabrication of Elastic Inter-connections for Stretchable Electronic Circuits. IEEE Electron. Device Letters 28(7), 552–554 (2007)
22. Torfs, T., Leonov, V., Vullers, R.J.M.: Pulse Oximeter Fully Powered by Human Body Heat. Sensors & Transducers Journal 80(6), 1230–1238 (2007)
23. Leonov, V., Gyselinckx, B., Van Hoof, C., Torfs, T., Yazicioglu, F., Vullers, R.J.M., Fiorini, P.: Wearable self-powered wireless devices with thermoelectric energy scavengers. In: Proceed. Conf. Smart Systems Integration, Barcelona, April 9-10, pp. 217–224 (2008)
24. Van Bavel, M., Leonov, V., Yazicioglu, R.F., Torfs, T., Van Hoof, C., Posthuma, N., Vullers, R.J.M.: Wearable Battery-Free Wireless 2-Channel EEG Systems Powered by Energy Scavengers. Sensors & Transducers 94(7), 103–115 (2008)

Experimental Assessment of Accuracy of Automated Knowledge Capture

Susan M. Stevens, J. Chris Forsythe, Robert G. Abbott, and Charles J. Gieseler

Sandia National Laboratories, P.O. Box 5800, Albuquerque, NM 87185, USA
{smsteve,jcforsy,rgabbot,cjgiese}@sandia.gov

Abstract. The U.S. armed services are widely adopting simulation-based training, largely to reduce costs associated with live training. However simulation-based training still requires a high instructor-to-student ratio which is expensive. Intelligent tutoring systems target this need, but they are often associated with high costs for knowledge engineering and implementation. To reduce these costs, we are investigating the use of machine learning to produce models of expert behavior for automated student assessment. A key concern about the expert modeling approach is whether it can provide accurate assessments on complex tasks of real-world interest. This study evaluates of the accuracy of model-based assessments on a complex task. We trained employees at Sandia National Laboratories on a Navy simulator and then compared their simulation performance to the performance of experts using both automated and manual assessment. Results show that automated assessments were comparable to the manual assessments on three metrics.

Keywords: Automated assessment, Naval training systems, simulation-based training, intelligent tutoring systems.

1 Introduction

A significant cost in simulation-based training is the workload on human instructors to monitor student actions and provide corrective feedback. For example, the U.S. Navy trains Naval Flight Officers for the E2-Hawkeye aircraft using a high-fidelity Weapons Systems Trainer (E2 WST). Currently this requires a separate instructor to observe each student within the context of team performance and provide instruction based on observed misunderstandings, inefficient task execution, ineffective or inappropriate actions, etc. Individualized instruction contributes to high training costs. Intelligent tutoring systems target this need, but they are often associated with high costs for knowledge engineering and implementation. New technologies are required that assist instructors in providing individually-relevant instruction.

1.1 Simulation Training

Establishing the validity of automated assessments requires studies in a realistic training environment, rather than just a simple laboratory task. E2 operators are trained and tested on several different simulators ranging from a part-task computer-based

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 212–216, 2009.

training (CBT) system that runs on a single PC, to the high-end E2 WST system which faithfully replicates most aspects of E2 operations (ranging from the physical controls to system fault diagnosis and recovery) and requires a team of instructors of operators to conduct training. For this study we used the E2 Distributed Readiness Trainer (EDRT), a medium-fidelity trainer which presents students with the same mission software used on the E2 aircraft. Multiple instructors are needed to evaluate simulation training and sessions can last hours at a time. Automated assessment of E2 operator performance in these sessions would greatly reduce instructor workload and would increase overall efficiency.

1.2 AEMASE

Sandia National Laboratories has shown the feasibility of automated performance assessment tools such as the Sandia-developed Automated Expert Modeling and Student Evaluation (AEMASE) software. One technique employed by AEMASE is the grading of students performance by comparing their actions to a model of expert behavior. Models of expert behavior are derived by collecting sample data from simulator exercises or other means and then employing machine learning techniques to capture patterns of expert performance. During training, the student behavior is compared to the expert model to identify and target training to individual deficiencies. Another technique utilized by AEMASE is the grading of students performance by comparing their actions to models of good and/or poor student performance. Students with good and bad performance are identified and machine learning techniques are employed to construct models of these two types of performance in the same manner as expert performance. Student performance from other training sessions is then compared to these models to identify and target training to individual deficiencies. Both techniques avoid the costly and time-intensive process of manual knowledge elicitation and expert system implementation (Abbott, 2006).

In a pilot study, AEMASE achieved a high degree of agreement with a human grader (89%) in assessing tactical air engagement scenarios (Abbott, 2006). However, the 68 trials assessed utilized only four subjects under three different initial training scenarios and the range of correct behaviors was quite limited. The current study provides a more rigorous empirical evaluation of the accuracy of these assessments. User modeling, based on behavioral and/or physiological measures, will be a key component of technologies implementing augmented cognition tools for training.

Purpose of Study. Automated assessments, such as AEMASE, would be a helpful tool in assessing E2 operator performance in an EDRT. Using AEMASE, user models can be derived with data generated from students executing scenarios within a simulation trainer or on actual equipment platforms. We trained employees at Sandia National Laboratories on an EDRT and then assessed their simulation performance using both AEMASE and manual assessment.

2 Methods

2.1 Participants

Twelve employees from Sandia National Laboratories volunteered to participate in the experiment. The participants met certain required criteria for the experiment which reflected the requirements for an entry-level E2 Hawkeye operator. In addition, two former E2 Hawkeye operators participated in the experiment and served as subject matter experts (SME's).

2.2 Materials

Materials included an E2 Deployment Readiness Trainer (EDRT) simulator that was obtained from the Naval Air Systems Command's Manned Flight Simulator organization. The Joint Semi-Automated Forces (JSAF) simulation software was used to create and drive the training and testing scenarios. In addition, the Sandia-developed Automated Expert Modeling and Student Evaluation (AEMASE) software and the Command Distributed Mission Training System (CDMTS) software were used in the analyses of the data.

2.3 Procedure

The participants were recruited via an advertisement and those who responded positively and met the required criteria were included in the study. The participants were scheduled for an initial all-day training session in which a former E2 Hawkeye Naval Flight Officer provided a tutorial on E2 operations emphasizing the basic radar systems task that would be the subject of the experiment. The participants were also asked to sign an informed consent. After the initial training session, the participants were scheduled for seven additional training sessions. The participants were lead through the sessions in the same order. Once they had finished the training sessions, the participants completed two testing sessions. The participants completed the seven training and two testing sessions individually.

Training Sessions. The first five sessions consisted of additional training sessions designed to teach the participants the basic operations of the E2 radar system in depth on the EDRT. For each session, the experimenters first demonstrated the proximate operation(s) on the EDRT and then the participant was asked to perform the operation(s) in scaled down, yet realistic, simulations. Since all five of these sessions were for training purposes, the experimenters were available to answer questions. At the end of each training session, the participants filled out a questionnaire indicating their understanding of the operation(s) on the preceding training session. At the end of the fifth scenario, the participants completed a questionnaire assessing their knowledge of all of the operations learned in the training sessions.

Testing Sessions. The last two sessions were testing sessions in which the participants were assessed on their knowledge of the operations and tactics covered in the five training sessions. The participants completed these more difficult simulations without the help of the experimenters. At the end of each testing session, the participants were asked to complete a questionnaire which queried their confidence of their performance on the preceding testing scenario.

Metrics. Based on guidance from the SMEs, three metrics were developed which were used to grade the participants' performance on the testing sessions. These metrics included fleet protection, labeling of neutral entities and Combat Air Patrol

(CAP) Asset Management. These metrics were used in both the manual and automated assessments.

Fleet Protection. Participants were instructed to prevent non-friendly entities from nearing the carrier group. The amount of time the non-friendly entities spent too close to the carrier group was assessed.

Labeling neutral entities. Participants were instructed to promptly and appropriately label any neutral entity that appeared on the radar scope. The latency with which the participants took to label these entities was assessed.

Combat Air Patrol (CAP) Assessment. Participants were instructed to effectively manage their air assets as the battle space evolved during the scenario. This included reordering CAP stations so that the airspace would not be violated.

Manual Assessment. Two trained experimenters independently reviewed video recordings of each of the testing scenarios for all participants. The experimenters graded the participants' performance on the three metrics for the two testing scenarios. For each metric, the two experiments specified at least one instance of good and one instance poor student performance. These instances formed subsets of manual assessment data that was used in training the AEMASE system.

Automated Assessment. The participant performance on the two testing scenarios was assessed by AEMASE. AEMASE used the good and poor instances indentified by the two experimenters as base examples from which to assess participant performance.

3 Results

The manual assessments and the automated assessments were compared for each of the three metrics.

Fleet Protection

Manual assessment was based on the amount of time the non-friendly fighters spent too close to the carrier group. The interrater reliability between the two experimenters was 99%. The automated assessment used a proxy measure, which consisted of the distance between the carrier group and the closest non-friendly asset. The results indicate a 100% agreement between the automated and manual assessments in terms of identification of unsatisfactory student performance (i.e., those students whose non-friendly assets got closest to the carrier group).

Labeling of entities

Manual assessment was based on reviewing the timestamped recording of when the neutral entities were labeled. The interrater reliability between the two experimenters was 94%. The automated assessment was based on the analysis of network messages from the mission computer. The results indicate a 95% agreement between the automated and manual assessment for correct labeling of the neutral entities.

CAP Station Rotation

Manual assessment was based on the time and accuracy with which the CAP stations were reordered. The interrater reliability between the two experimenters was 99%. The automated assessment was based on post-hoc analysis of radio communications. Results indicate an 83% agreement between the automated and manual assessment.

4 Discussion

AEMASE surpassed target performance criteria with agreement of up to 100% with the manual assessment. Even with an undeniably difficult metric that was based on radio communication (the CAP station rotation metric), agreement between AEMASE and manual assessment was an impressive 83%.

Acknowledgements

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. This work was performed through a contract award from the Office of Naval Research.

Reference

1. Abbott, R.G.: Automated expert modeling for automated student evaluation. Intelligent Tutoring Systems, 1–10 (2006)

Eye Movement as Indicators of Mental Workload to Trigger Adaptive Automation

Tjerk de Greef[1,3], Harmen Lafeber[2], Herre van Oostendorp[2], and Jasper Lindenberg[3]

[1] Delft University of Technology, Man-Machine Interaction Group, Mekelweg 4, 2628 GA Delft, The Netherlands
[2] University of Utrecht, Department of Information and Computing Science, Padualaan 14, 3594 CH Utrecht, The Netherlands
[3] TNO Defence, Security and Safety, Department of Human Factors, Kampweg 5, 3769 ZG Soesterberg, The Netherlands
t.e.degreef@tudelft.nl, h.lafeber@gmail.com, herre@cs.uu.nl, jasper.lindenberg@tno.nl

Abstract. This research describes an approach to objective assessment of mental workload, by analyzing differences in pupil diameter and several aspects of eye movement (fixation time, saccade distance, and saccade speed) under different levels of mental workload. In an experiment, these aspects were measured by an eye-tracking device to examine whether these are indeed indicators for mental workload. Pupil diameter and fixation time both show a general significant increase if the mental workload increases while saccade distance and saccade speed do not show any significant differences. This assessment of mental workload could be a trigger for aiding the operator of an information system, in order to meet operational requirements.

Keywords: mental workload, adaptive automation, eye movement, pupil diameter, saccade, fixation time.

1 Introduction

In 1988, in the straight of Hormuz, the USS Vincennes mistakenly shot down a commercial Iranian airplane because it was misidentified as an Iranian F-14 combat airplane. Much is written [1] about contributing factors that lead to this tragic accident and some authors summarize it as human error. This is a typical example of a high-risk professional domain where humans carry large responsibility and where mistakes result in tragic accidents and/or heavy losses. In these information-rich and dynamic environments, a competition for the human's attention is going on between numerous different information items, at times leading to a cognitive overload. This overload originates from the limitations in human attention and constitutes a well-studied bottleneck in human information processing. If the human is getting overloaded, a control mechanism capable of adjusting the balance of work between the human operator and the machine might lower the cognitive burden of the human and in effect optimize the performance of the human machine ensemble. A so-called adaptive system [2] in which the division of labor between human and machine is flexible and

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 219–228, 2009.

responsive to task or human demands, is thought to represent a better solution to the problem of function allocation than the static ones currently in use. Rouse [3] introduced adaptive aiding as an initial type of adaptive automation and stated that adaptive aiding is a human-machine system-design concept that involves using aiding/automation only at those points in time when human performance needs support to meet operational requirements [3]. Whether one uses the terms adaptive automation, dynamic task allocation, dynamic function allocation, or adaptive aiding, they all reflect the dynamic reallocation of work in order to improve human performance or to prevent performance degradation.

As a matter of fact, adaptive automation should scale itself down when things become quieter again and the goal of adaptive automation could be stated as trying to keep the human occupied within a band of 'proper' workload [see 4]. Periods of 'underload' can have equally disastrous consequences as periods of overload due to slipping of attention and loss of situational awareness.

As stated, an adaptive system changes the division of labor between the human and machine as the effectiveness of a human operator is a concern in relation to the task demands. Parasuraman [5], for example, found superior performance when the control of a fault management task (i.e., to monitor an automated system and diagnose the problem in case the system halts) was allocated back to the human for some time. Other studies shift control from the human to the machine in case the human is incapable or indecisive to make a decision as seen in the (no)go decision in case of a engine failure in the takeoff run of an airplane [6]. More recently, adaptive automation is applied to the domain of naval warfare [7] where part of the identification of airplanes or vessels is executed by the system when the human starts to fall behind. In synopsis, a number of studies have shown that the application of adaptive automation enhances performance, reduces workload, improves situational awareness, and maintains skills that are deteriorating as a consequence of too highly automated systems [5-9].

One of the challenging factors in the development of successful adaptive automation concerns the question of when changes in the level of automation must be effectuated. Wilson and Russell [10] define operator psychophysiology as one of five triggering strategies based on an previous division by Parasuraman *et al* [5]. Psychophysiological data from the operator are employed in various studies [9-13] and prove an objective measure. Examples of these measurements are: heartbeat rate, respiratory, facial expressions, perspiration, eye blink rate (see [14] for an overview). Although various studies [9-13] indicate a mental workload effect on psychophysiological characteristics, no single psychophysiological measure can be directly interpreted as such [15]. Variations in psychophysiological measurements, however, can be assigned to a lot of different aspects, with mental workload just being one of them. The main advantages of objective measurements are that they do not have to interrupt the operator in task execution.

One popular type of psychophysiological data measurement involves workload effects on properties and movement of the eye [16-20]. Although a number of studies found [17-19] empirical evidence in the favor of utilizing an increased pupil diameter as an of increased mental workload, not all studies have obtained similar results. Kramer [21], for example, relates the failure to find similar results to factors unrelated to the task that produce larger changes in pupil dilatation such as changes in ambient illumination or screen luminance.

Alternatively, other relations between mental workload and properties of the eye exist. During visual scanning, muscles direct the eye to interesting areas where fixations occur. A fixation is usually defined as a steady focus of the eye for 100 to 200 milliseconds, which provides the visual system with detailed input about the visual stimulus. Simultaneously, parallel processes use peripheral visual information to determine where the next fixation will be located [22]. The movement to another fixation stimulus is defined as a saccade. Tole, Harris, Stephens, and Ephrath [20] found an increase in fixation time when the mental workload increased. However, saccade measurements show no consequent results in relation to mental workload [16, 18].

The previous paragraphs clearly indicate a challenge in utilizing eye properties as indicator for mental workload. One side of the scientific literature shows that pupil diameter, fixation times, saccade distance, and saccade speed can be used as an indicator of mental workload while other studies show counterarguments. We are interested how the properties of the eye respond in various workload conditions in a naval warfare domain. Once successful, these properties of the eye can be used to trigger adaptive automation.

We conducted an experiment where certain properties of eye movement and pupil diameter were measured under different levels of mental workload. Consequently we question whether fixation time, saccade distance, saccade speed, and pupil diameter can be used as objective indicator for mental workload in such a task setting.

As the study evolves around a measure of mental workload, we will manipulate mental workload using a validated model of cognitive task load. Therefore our first hypothesis reads that:

1: three scenarios are generated having a predicted and different mental workload.

Using these differences in mental workload we can perform measurements on properties of the eye. Following experimental effects on pupil diameter and fixation times found in respectively [17-19] and [20], we hypothesize that:

2: if the mental workload of an operator increases, pupil diameter increases, and
3: if the mental workload of an operator increases, fixation time increases.

Furthermore, it is expected that saccade distance will decrease in response to an increase of mental workload due to an effect called tunnel vision (i.e. the loss of peripheral vision with retention of central vision, resulting in a constricted circular tunnel-like field of vision). Also, saccade speed is expected to decrease due to fatigue of the muscles, as evidence for fatigue in pupillary muscles exists [23]. Consequently, we hypothesize that

4: if the mental workload of an operator increases, saccade speed decreases, and
5: if the mental workload of an operator increases, saccade distance decreases.

The next section discusses the method & materials used in this study and section 3 presents the results of the experiment. Section 4 discusses the results and draws conclusions from these results from the perspective of adaptive automation.

2 Research Method

2.1 Participants

Eighteen subjects participated in the experiment and were paid to participate. The test subjects were all university students, with a good knowledge of English. The participant group consisted of ten men and eight women. They had an average age of 25, with a standard deviation of 5.1.

2.2 Experimental Tasks

The subjects were given the role of human operators of (an abstracted version of) a combat management workstation (CMS) aboard naval vessels. The workstation comprised a schematic visual overview of the nearby area of the ship on a computer display, constructed from the data of radar systems. On the workstation the subject could manage all the actions required to achieve mission goals.

More specifically, the goal of the human operator during the scenarios was to monitor, classify, and identify every track (i.e. airplanes and vessels) within a 38 nautical miles range around the ship. Furthermore, in case one of these tracks showed hostile intent (in this simplified case a dive toward the ship), they were mandated to protect the naval vessel and eliminate the track.

To achieve these goals, the subject was required to perform three tasks. First, the classification task gained knowledge of the type of the track and its properties using information from radar and communication with the track, air controller, and/or the coastguard. The subject could communicate with these entities using chat functionality within the CMS. The experiment leader responded to such communications. The second task was the identification process that labeled a track as friendly, neutral, or hostile. The last task was weapon engagement in case of hostile intent as derived from certain behavior. The subject was required to follow a specific procedure to use the weapons.

2.3 Scenarios

We designed three different scenarios, each implying a different cognitive task load. The task loads were under-load, normal load, and an overload achieved by manipulating two of the three cognitive task load factors as defined in Neerincx model [24] of cognitive task load (CTL).

The CTL model is comprised of three factors that have a substantial effect on the cognitive task load. The first factor, percentage time occupied, has been used to assess workload for time-line assessments. The second load factor is the level of information processing that addresses cognitive task demands. The model therefore incorporates the skill-rule-knowledge framework of Rasmussen [26] where the knowledge-based component involves the highest workload. To address the demands of attention shifts, the model distinguishes task-set switching as a third load factor. It represents the fact that a human operator requires time and effort to reorient himself to a different context. These factors present a three-dimensional space in which all human activities can be projected as a combined factor. Specific regions indicate the cognitive demands activities impose on a human operator.

Creating scenarios using the CTL mode has been applied successfully in a number of experimental [27] and realistic [28] settings. We applied the model to implement a certain cognitive load. First, the total number of tracks in a scenario was changed. If many tracks are in the observation range, the percentage of the total time that the human is occupied is high. Second, a larger amount of tracks that show special behavior and more ambiguous properties increases the operator's cognitive workload due to applying more rule and knowledge based reasoning. It forces the human operator to focus attention and to communicate more in order to complete the tasks.

2.4 Variables and Experimental Design

In order to control for intra-individual variability in cognition, we chose to use a within-subject design. In order to limit the potential for individual differences on any experimental condition and to filter out sequence effects, we applied a Latin square design to the combination of independent variables and the sequence of scenarios.

The independent variable was the workload as manipulated in a scenario (see previous section). Furthermore, five dependent variables were measured:

- Mental workload was measured and controlled using an adapted version of a workload watch [25] that signaled the subject every 100 seconds to rate his/her perceived workload on a scale (one to five), by clicking on the corresponding button present on the lower right of the screen. Button 1 indicated low workload, button 3 normal workload and button 5 high workload. The buttons in between indicate intermediate levels of workload.
- Pupil diameter in micrometers during the trials (averaging both eyes).
- Fixation time: the time that fixations lasted within a radius of 40 pixels and a minimum of 100 milliseconds. The fixation times were divided by the total number of fixation to derive the average time a fixation lasts and only the fixations on a track were accounted to get a reliable representation of cognitive processing time.
- Saccade distance: the distance in pixels between one fixation and the next.
- Saccade speed defined as the saccade distance divided by the saccade time.

2.5 Apparatus

The CMS application was run on a computer connected to a 17-inch monitor, with a resolution 1280x1024. The eye-tracking device Tobii X50 was connected to another computer. The experimental leader was situated behind a desk with a third computer running the same scenario as the test person to ensure good communication.

The Tobii X50 recorded the required aspects of eye movement and pupil diameter of the subjects during the task. The Tobii X50 was placed in front and underneath the monitor that the subjects used for the task. This was approximately 60 cm in front of them.

2.6 Procedure

Before the experiment, the subjects were given a clear description of the various tasks to be executed during the scenarios and various test round were offered to the subjects. Before every scenario, a description about the position of the naval ship and its mission was provided. The experiment was conducted in a closed room where the subjects were not disturbed during the task. During the experiment, an experimental leader was situated roughly two meters behind the subject to assist when necessary.

3 Results

For each dependent variable a repeated-measures analysis ANOVA with within-factor scenario was used to analyze the data as the subjects are exposed to each condition in turn. In all cases, an alpha level of .05 was used to determine statistical significance. The data were analyzed using SPSS. For post-hoc analysis, the least significant difference (LSD) test was used and the partial eta square statistics (η_p^2) was adopted to describe the estimated proportion of variance explained by the factors. The partial eta square has the advantage that it is independent on the number of factors.

Verification of the mental workload used data of all but one subject due to a failure in logging (N = 17). The rest of the statistical tests utilized N = 13 because the eye-tracker data from four subjects could not be used due to a technical failure of the eye-tracking device in one of the three scenarios.

3.1 Workload Verification

Repeated-measures ANOVA reveals a significant effect in subjective (indicated) mental workload between the three scenario's ($F(2, 33) = 190.632$, $p < .001$, $\eta_p^2 = .923$). Least square difference post-hoc analysis reveals that all three means were significantly different ($p < .05$). Compared to the under-load scenario, the perceived mental workload was significantly higher in the normal workload scenario. In turn, the perceived mental workload in the overload scenario was significantly higher again than in the normal-workload scenario (see Fig. 1).

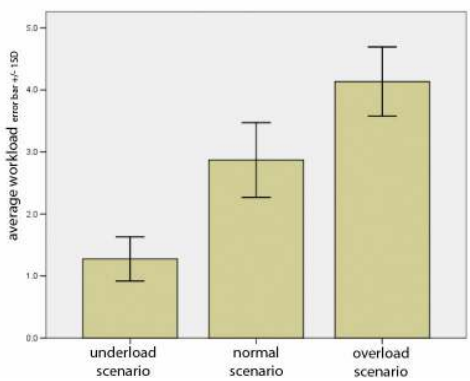


Fig. 1. The subjective workload per scenario as indicated every 100 seconds on a five point scale

3.2 Pupil Diameter

Repeated measures one-way ANOVA reveals that there are significant differences in pupil diameter between the three workload scenario's, $F(2, 69) = 3.720$, $p < .005$, $\eta_p^2 = .237$). A post-hoc LSD comparisons revealed that the underload scenario and the overload scenario were significantly different. Compared to the underload scenario ($M = 4957.42$, $SD = 511.164$), the average pupil diameter was significantly higher in the overload scenario ($M = 5094.90$, $SD = 547.797$). Furthermore the normal scenario ($M = 5069.84$, $SD = 152.676$) had a significant higher average pupil diameter in comparison with the underload scenario. However, no significant differences were found between normal scenario and the overload scenario.

3.3 Fixation Time

SPSS repeated measures one-way ANOVA displays a significant effect of fixation time on the different types of scenario ($F(2, 69) = 13.411$, $p < .05$, $\eta_p^2 = .528$). LSD comparisons revealed that all scenario's were significantly different from each other. Compared to the underload scenario ($M = 249.177$, $SD = 113.192$), the average fixation time was significantly higher in the normal scenario ($M = 287.992$, $SD = 108.710$) and on its turn the normal scenario differed significantly from the overload scenario ($M = 334.838$, $SD = 130.823$).

3.4 Saccade Distance and Speed

No significant differences are found with a repeated measures one-way ANOVA in saccade distance between the three scenario's ($F(2, 69)= 0.388$, $p = .683$, $\eta_p^2 = .031$). In addition, no significant differences are found with a repeated measures one-way ANOVA in saccade speed between the three scenario's ($F(2,69) = 1.768$, $p = .192$, $\eta_p^2 = .128$).

4 Conclusion and Discussion

In high-risk professional domains, like a naval warfare, it is preferred to keep humans in control to handle unanticipated novel situations. However, increased mental workload might decrease human performance resulting in endangering mission goals. Therefore it is important to detect such situations. Various studies suggest that the machine should aid the human operator in those situations to meet operational requirements [3]. However the measurement of such situations is much debated and this research describes an approach to assess mental workload objectively by analyzing differences in pupil diameter, fixation time, saccade distance, and saccade speed under different levels of mental workload. We consequently conducted an experiment to measure these four aspects under various workload conditions to ascertain their utility as an objective mental workload indicator.

The results show that the manipulation of the scenario's worked as expected in that the manipulation of two of the three CTL model variables resulted in significant different subjective mental workload. The results not only confirm hypothesis 1 but also extend the knowledge [27, 28] on the successful application of the CTL model to design scenarios with an intended mental workload.

Although the results reveal a general significant effect of pupil diameter on workload manipulations, the results fail to discriminate between all workload manipulations. We therefore partly accept hypothesis 2 because of the general nature to utilize pupil diameter as an indicator of workload. Consequently we agree with [21] that the pupil diameter responds to many factors with workload being one of them in contrasts other research that found pupil diameter effects [17-19].

Results show, on the other hand, a significant and discriminatory effect of mental workload manipulations on fixation time. This means that we accept hypothesis 3 because fixation time can be used to distinguish between workload conditions. The results comply with research by Tole [20] who found an increase in fixation time when the mental workload increased and [29] found a negative correlation between fixation time and performance. This complies with the findings in this research, given that higher fixation times indicate a higher level of mental workload and mental workload has a negative effect on performance.

Saccade distance and saccade speed show no significant differences when the mental workload increases and we therefore reject hypothesis 4 & 5. These results comply with research [16] that failed to find a relation between saccade measures and mental workload. However, [18] did find a decrease in saccade distance if mental workload increases but stated that many properties of the eye, including saccades, are highly task dependent.

Much research has been done to find a relation between operator psychophysiology and operator workload. As stated before, no single psychophysiological measure can be interpreted as a workload indicator. Therefore, in order to obtain a reliable objective indicator for mental workload, it is necessary to work towards a model that integrates several psychophysiological measurements. The construction of such a model is complex because the reliability and relative importance of the different measurements are hard to define. The results of this research contribute to this model by examining the effects on pupil diameter, fixation time, saccade distance, and saccade speed under different levels of mental workload in a naval warfare setting.

As this research shows that measurements of properties of the eye with an eye-tracking device can provide a valuable addition to the determination of the level of aiding, problems arise when it comes to the practical application of the concept. As indicated throughout the paper, many factors influence workload and properties of the eye. In an experiment, these aspects can be kept as constant as possible. In a practical application, for example with defense tasks on a navy ship, it cannot be expected of the human operator to refrain from drinking caffeine-holding beverages. These aspects make the deduction of mental workload from for example pupil diameter unreliable. However, if a combination of psychophysiological measurements is used and they all indicate a similar operator workload, this indication can be very usable.

Acknowledgements

This study was supported by the MOD-NL DR&D organization under program 'human system task integration' (V206). The department of Human in Command of TNO Defence, Security and Safety kindly provided an internship fee.

References

1. Cannon-Bowers, J.A., Salas, E.: Making decisions under stress: implications for individual and team training. American Psychological Association, Washington (1998)
2. Scerbo, M.: Theoretical perspectives on adaptive automation. In: Parasuraman, R., Mouloua, M. (eds.) Automation and human performance: theory and applications, pp. 37–63. Lawrence Erlbaum Associated Publishers, Mahwah (1996)
3. Rouse, W.B.: Adaptive Aiding for Human/Computer Control. Human Factors 30, 431–443 (1988)
4. Endsley, M., Kiris, E.: The Out-of-the-Loop Performance Problem and Level of Control in Automation. Human Factors, 381–394 (1995)
5. Parasuraman, R., Mouloua, M., Molloy, R.: Effects of adaptive task allocation on monitoring of automated systems. Human Factors 38, 665–679 (1996)
6. Inagaki, T.: Situation-adaptive autonomy for time-critical takeoff decisions. International Journal of Modelling and Simulation 20, 175–180 (2000)
7. Arciszewski, H.F.R., de Greef, T.E., van Delft, J.H.: Adaptive Automation in a Naval Combat Management System. IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans (in press)
8. Moray, N., Inagaki, T., Itoh, M.: Adaptive Automation, Trust, and Self-Confidence in Fault Management of Time-Critical Tasks. Journal of experimental psychology, 44–57 (2000)
9. Bailey, N.R., Scerbo, M.W., Freeman, F.G., Mikulka, P.J., Scott, L.A.: Comparison of a brain-based adaptive system and a manual adaptable system for invoking automation. Human Factors 48, 693–709 (2006)
10. Wilson, G.F., Russell, C.A.: Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding. Human Factors 49, 1005–1018 (2007)
11. Byrne, E.A., Parasuraman, R.: Psychophysiology and adaptive automation. Biological Psychology 42, 249–268 (1996)
12. Veltman, J.A., Gaillard, A.W.K.: Physiological workload reactions to increasing levels of task difficulty. Ergonomics 41, 656–669 (1998)
13. Prinzel, L.J., Freeman, F.G., Scerbo, M.W., Mikulka, P.J., Pope, A.T.: A closed-loop system for examining psychophysiological measures for adaptive task allocation. International Journal of Aviation Psychology 10, 393–410 (2000)
14. Weert, J.C.M.: Ship operator workload assessment tool. Department of mathematics and computer science. Technical University Eindhoven, Eindhoven (2006)
15. Veltman, J.A., Gaillard, A.W.D.: Pilot workload evaluated with subjective and physiological measures. In: Brookhuis, K.A., Weikert, C., Moraal, J., de Waard, D. (eds.) Proceedings of the European Chapter of the Human Factors and Ergonomics Society, Soesterberg, the Netherlands (1996)
16. Brookings, J.B., Wilson, G.F., Swain, C.R.: Psychophysiological responses to changes in workload during simulated air traffic control. Biological Psychology 42, 361–377 (1996)
17. Iqbal, S.T., Zheng, X.S., Bailey, B.P.: Task-evoked pupillary response to mental workload in human-computer interaction. In: Proceedings of the ACM Conference on Human Factors in Computing Systems, Vienna, Austria, pp. 1477–1480 (2004)
18. van Orden, K.F., Limbert, W., Makeig, S., Jung, T.: Eye activity correlates of workload during a visuospatial memory task. Human Factors 43, 111–121 (2001)
19. Beatty, J.: Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychological Bulletin 91, 276–292 (1982)
20. Tole, J.R., Harris, R.L., Stephens, A.T., Ephrath, A.R.: Visual scanning behavior and mental workload in aircraft pilots. Aviation, Space, and Environmental Medicine 53, 54–61 (1982)

21. Kramer, A.F.: Physiological metrics of mental workload: A review of recent progress. In: Damos, D.L. (ed.) Multiple-task performance, Taylor & Francis, London (1991)
22. Henderson, J.M.: Visual attention and eye movement control during reading and picture viewing. In: Rayner, K. (ed.) Eye movements and visual cognition - Scene perception and reading, pp. 260–283. Springer, New York (1992)
23. Stern, J.A.: The pupil of the eye: what can it tell about mental processes? Human engineering for quality and life 8, 1–2 (1997)
24. Neerincx, M.A.: Cognitive task load design: model, methods and examples. In: Hollnagel, E. (ed.) Handbook of Cognitive Task Design, pp. 283–305. Lawrence Erlbaum Associates, Mahwah (2003)
25. Boer, L.C.: Workload-watch as an element of human engineering testing and Evaluation. In: Eleventh ship control systems symposium, vol. 2, Computational mechanics publications, Southampton, United Kindom (1997)
26. Rasmussen, J.: Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering. North-Holland, Amsterdam (1986)
27. Neerincx, M.A., Besouw, N.J.P.: Cognitive task load: a function of time occupied, level of information processing and task-set switches. Industrial Ergonomics, HCI, and Applied Cognitive Psychology 6, 247–254 (2001)
28. Grootjen, M., Neerincx, M.A., Veltman, J.A.: Cognitive task load in a naval ship control centre: From identification to prediction. Ergonomics 49, 1238–1264 (2006)
29. Goldberg, J.H., Kotval, X.P.: Eye movement-based evaluation of the computer interface. In: Kumar, S. (ed.) Advances in occupational ergonomics and safety, pp. 529–532. ISO press, Amsterdam (1998)

Impact of Automation and Task Load on Unmanned System Operator's Eye Movement Patterns

Cali M. Fidopiastis[1], Julie Drexler[1], Daniel Barber[1], Keryl Cosenzo[2], Michael Barnes[2], Jessie Y.C. Chen[2], and Denise Nicholson[1]

[1] Applied Cogntion and Training in Immersive Virtual Enviroments (ACTIVE) Laboratory
Institute of Simulation and Training (IST), University of Central Florida
[2] U.S. Army Research Laboratory (ARL) - Human Research & Engineering Directorate
{cfidopia,jdrexler,dbarber}@ist.ucf.edu, kcosenzo@arl.army.mil,
{michael.j.barnes,jessie.chen}@us.army.mil, dnichols@ist.ucf.edu

Abstract. Eye tracking under naturalistic viewing conditions may provide a means to assess operator workload in an unobtrusive manner. Specifically, we explore the use of a nearest neighbor index of workload calculated using eye fixation patterns obtained from operators navigating an unmanned ground vehicle under different task loads and levels of automation. Results showed that fixation patterns map to the operator's experimental condition suggesting that systematic eye movements may characterize each task. Further, different methods of calculating the workload index are highly correlated, $r(46) = .94$, $p = .01$. While the eye movement workload index matches operator reports of workload based on the NASA TLX, the metric fails on some instances. Interestingly, these departure points may relate to the operator's perceived attentional control score. We discuss these results in relation to automation triggers for unmanned systems.

Keywords: Adaptive Automation, Unmanned Ground Systems, Eye Tracking, Workload.

1 Introduction

Unmanned air and ground vehicles are the inevitable future of the United States Army Future Combat Systems [1]. Suggested roles for unmanned ground vehicles (UGVs), in particular, are as remote weapons platforms, soldier companions, or substitutes for reconnaissance, surveillance, and target acquisition (RSTA). The level of autonomy, or capability of the UGV to perform tasks without human assistance, is determined by the robot's primary function [2]. For example, teleoperated ground vehicles are not typically equipped with onboard terrain or environment maneuvering capabilities; consequently, their level of autonomy is low. The human assumes the role of an operator, monitoring the environment and making all task related decisions.

In contrast, semi-automated vehicles are intelligent enough to assume a navigator role with point A to B mobility. Thus, the human team member can perform other tasks requiring higher cognitive skills (e.g., problem solving) that the robot is not capable of, or less successful at, executing. The logical compromise between human

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 229–238, 2009.

and machine systems such as these is to provide a means to share or trade control over tasking through adaptive automation [3] [4].

Automation within human-machine systems can be thought of as a tool that allows the human to turnover designated tasking to a machine agent, thereby improving overall work performance [5]. Adaptive automation (AA) refers to a system capability that enables flexible task allocations between the human operator and the machine agent in the context of the work environment [6]. Within high stress, time-pressured operations, the goal of the adaptive system would be to reduce operator workload and fatigue by automating select activities. In contrast, during times of underload, the system would reengage the operator to improve situational awareness by returning tasks. Thus, the human-machine system trades off assignments cooperatively, yet with an emphasis on the human operator. This human-centered approach keeps the operator in the control loop, which minimizes deleterious performance issues such as reorienting to tasks slowly and overlooking automation failures [7].

Developers of AA use several strategies to determine when to invoke changes in automation level. These strategies include: 1) operator performance measures, monitoring operator performance criterion levels; 2) critical events, signaling system failures or environmental changes; 3) system models, comparing human-machine performance to an *a priori* determined pattern of system performance; 4) operator state assessments, using psychophysiological measures to determine operator workload; and 5) hybrid methods, combining one or more of the above strategies [7] [5] [8]. With the exception of hybrid methods, assessing the operator's state using psychophysiological measures is the only triggering technique that provides real-time monitoring of the operator's state.

Monitoring the operator's psychophysiological changes in real-time additionally provides an unobtrusive means to evaluate his or her internal status without interrupting the task. Given that the human operator may be construed as an adaptive system that can rapidly shift mental and physical resources or change cognitive strategies to meet task demands, identifying detrimental system shifts such as high mental workload at onset, or even prior to onset, may be paramount to successful human-robotic team efforts [6]. For example, cognitive workload levels are typically higher when a human operator teleoperates a UGV or intervenes when an autonomous operation fails [9]. Increased task load concomitant with increases in cognitive workload lead to reduced situational awareness [7] [10]. These effects increase the potential for catastrophic errors that can harm not only the team, but compromise the mission as a whole.

Task load, however, is not the only determinant of degraded human performance within human-robotic teams. As Chen and Terrence (2008) showed, operator's perceived attentional control (PAC) scores were associated with error rates determined by different types of unreliable automation. More specifically, persons scoring high on attentional control were more affected by false-alarm prone alerts than miss-prone alerts. A relationship between two types of aided target recognition (AiTR), tactile or visual cueing, and spatial ability scores of the participants was also found. Persons with low spatial ability preferred visual cueing, while person's with high spatial ability scores preferred tactile cueing. These examples of individual differences imply that hybrid automation strategies may provide a more accurate means of varying automation during real-world tasking. Further, these strategies should include real-time operator state assessments [11] [4].

The focus of the current research is to determine automation timing through objective psychophysiological measures. In this paper, we present preliminary data on assessing operator workload using eye gaze patterns and operator characteristics such as attentional control. These results extend previously reported work [12].

Chen et al. (2008) simulated a generic mounted crew station and examined the workload and performance of a UGV operator. The operator completed one of four types of tasking: 1) navigating the robot only, 2) navigating the robot plus a visual monitoring task, 3) navigating the robot plus an auditory monitoring task, and 4) navigating the robot plus both the auditory and visual tasks. These tasks were performed with differing levels of robot autonomy (teleoperated versus semi-autonomous) and levels of aiding for target recognition (AiTR versus no AiTR). While the psychophysiological measure of heart rate variability failed to predict changes in workload with changing task load, other findings support Chen and Terrence (2008). For example, participants with low PAC scores performed worse when teleoperating the robot than did those with higher PAC scores.

Participants' eye tracking data collected during the Chen et al. (2008) experiment was analyzed for this study. It was hypothesized that gaze patterns as assessed by fixations within areas of interest (AOI) could predict the type of tasking condition experienced by the operator [13]. Further, these visual patterns may predict the mental workload level of the operator [14]. Discussed here is the assessment of the utility of the eye fixation-based workload metric in future studies of adaptive automation.

2 Method

2.1 Participants

A total of 64 students from the University of Central Florida and the U. S. Military Academy participated in the original study. Eye tracking data from nine participants (Female = 3 and Male = 6) were further analyzed for this current study based on participant's high task performance scores and the existence of a complete eye tracking dataset. Participants ages ranged from 18 to 25 (M = 20.44, SD = 2.60).

2.2 Apparatus

2.2.1 Simulation

The Mixed Initiative eXperimental (MIX) testbed is a distributed simulation environment designed to empirically evaluate human-robot interactions and the affects of automation on operator performance. Details of the design and capabilities of the MIX testbed are provided in Barber, Davis, Nicholson, Finkelstein, and Chen [15]. We outline the robotic vehicle control capabilities, along with the operator control interface.

Robotic Vehicle Control. The MIX testbed can generate simulated unmanned ground or air vehicles through its Unmanned Vehicle Simulator (UVSIM). A UGV was used in this experiment. The UGV system supports manual control and two types of automation: waypoint navigation (i.e., semi-autonomous) and AiTR. In teleoperation mode, the operator must manually navigate the UGV along the mission route through the use of a standard joystick. Waypoint navigation allows the UGV to follow a preplanned route without operator input. There are four checkpoints encountered along the mission route where the operator must pause and perform a RSTA (Reconnaissance, Surveillance, and Target Acquisition) task. In this experiment, participants scanned for friendly versus foe targets either with AiTR support or without. When enabled, AiTR pans the vehicle's camera (up to 360 degrees), scans for targets within the environment, and generates a list of all targets (friendly and foe) within the scene. Without this automation, the operator must perform the target acquisition procedure him or herself through the interface controls. Operators performed each of the four control conditions, teleoperation with and without AiTR and semi-autonomous with andwithout AiTR, under varying task load.

Operator Control Unit. The Operator Control Unit (OCU) provides the graphical user interface with which to interact with the UGV. For example, Figure 1 shows the button controls along with the main viewing screen (center) and a "bird's eye" view of the UGV location on the upper left. The navigation bar allows the operator to zoom in/out as well as move across the maps. In addition, auditory and visual monitoring tasks are available to increase the operator's task load and simulate real-world conditions.

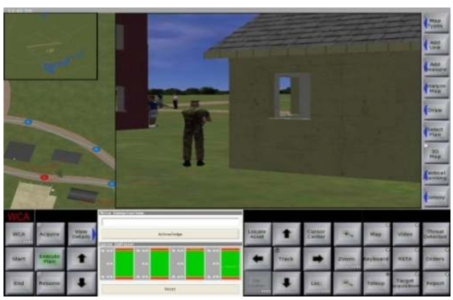


Fig. 1. OCU areas used for AOI for the fixation analysis

The visual monitoring display is also shown in Figure 1. This task required the operator to monitor the four gauges and assess whether they entered critical high or low levels. When these critical levels were detected, the operator pressed a "reset' button on the OCU display to return them to normal. The auditory task required the operator to monitor a series of auditory cues (i.e., call signs) periodically presented throughout the scenario and type his or her responses to each cue in the communication panel located below the visual monitoring display. These three areas, main view, bird's eye view, and secondary task display, were also used as the areas of interest (AOI) for the eye fixation analysis, which we describe next.

2.2.2 Eye Tracker and Nearest Neighbor Index

Figure 2a shows the Applied Science Laboratories (ASL) R6000 remote pan/tilt head and monocular eye tracker system that was used to track both head and eye movements during the study. The head and eye tracker system uses facial features, as shown in Figure 2b, to first detect the position of both eyes. The eye camera and eye illuminator are integrated into the pan and tilt mechanism to allow the system to dynamically move as the operator changes head position. Using a bright pupil technique, the illuminator projects near infrared light through the pupil such that the light reflects off the retina. The reflection created by the returning light traversing the cornea produces a bright pupil effect. This allows the eye camera, which captures pictures of the eye every 17 ms (60 Hz), to estimate the eye pupil center and to discriminate the corneal reflection (CR). These two parameters are then used to estimate the point of gaze of the operator.

Because the relationship between the pupil center and the CR are different for everyone, each participant performed a calibration procedure before entering the experiment. Participants were directed to look at 17 points (i.e., pre-defined XY coordinates) one at a time at different locations on the display. The eye tracker then mapped the participants' eye position relative to the monitor used to display the OCU. The accuracy of the point of gaze estimation by the ASL eye tracker is less than 1 degree visual angle (roughly a 7 mm error when the participant is seated at 76 cm from the display, as in this study).

Eye movement during viewing allows the fovea of the retina to scan an area, providing the fine details necessary to accurately locate and identify objects within a scene [16]. In general, rapid eye movements (velocities reaching 500 degrees per second) are typically referred to as saccades and provide little new information about the changing visual information. Fixations, which are brief moments of stability (roughly 200-300 ms and depend on the task), facilitate visual processing [17]. Feature detection algorithms are needed to identify saccades and fixations within the stream of raw eye tracking data. Given the type of eye tracker and the type of task,

Fig. 2a.

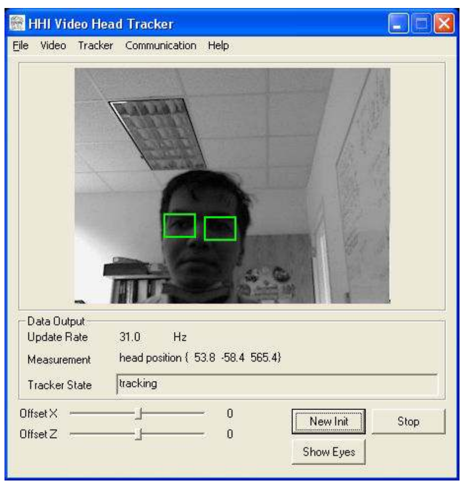


Fig. 2b.

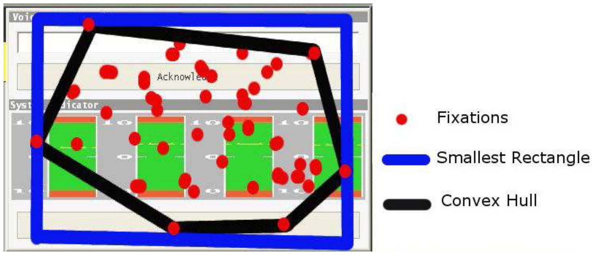


Fig. 3. Smallest rectangle and convex hull area methods for calculating the NNI

fixations and saccades were calculated using a dispersion-based algorithm that accounts for both the temporal and the spatial characteristics of eye movements [18].

Eye movements are driven by task requirements; current research suggests that only objects needed for the task are fixated upon in a serial manner when performing everyday activities [13]. Workload may be indirectly measured through the dispersion pattern or distribution of fixations in regions of the scene important for executing a task [19]. More specifically, persons experiencing periods of high workload may perform more stereotyped scanning patterns across the work space. The Nearest Neighbor Index (NNI) was developed to capture shifts between random and more focused fixations during real-world task performance and relate these shifts to levels of workload [14]. The goal of this effort was to develop an unobtrusive and reliable metric for applying adaptive automation within operational environments.

The NNI is calculated as a ratio, where the numerator represents the average nearest neighbor distances for fixation points and denominator represents the mean random distances for an AOI that one would expect if the fixation distribution were random [19]. Thus, ratio values closer to 1 suggest a random distribution of fixations or low workload, while the converse is true for high workload. The main and bird's eye views along with the visual and the auditory monitoring task display were designated as the AOIs for this study. Figure 3 shows two methods for determining clusters of fixation points in an AOI, smallest rectangle and convex hull (i.e., convex polygon). NNI's were calculated using both methods for comparison.

2.2.3 Attentional Control and NASA TLX Questionnaires

Perceived attentional control (PAC) was measured using the attentional control survey [20]. The survey consisted of 21 items and provided a measure of perceived attention focus and shifting. Research shows a relationship between eye fixations and focused attention [16]. Given that the scale has good internal reliability ($\alpha = .88$), it may provide a means to further assess the NNI metric.

The NNI has also been assessed against the National Aeronautics and Space Administration-task load index [2] [19]. The NASA-TLX is a self-reported questionnaire of perceived demands in six areas: mental, physical, temporal, effort (mental and physical), frustration, and performance [21]. The computerized version was used in the original study. We also compare the NNI with the NASA-TLX scores.

2.2.4 Experimental Design

NNI scores were only calculated for participants who had a complete eye tracking data and performed the best within their task load condition and level of automation (e.g., robotic x teleop x no AiTR). A participant could represent more than one condition because of missing data. However, given that PAC is a global measure of attentional control and the NASA-TLX was reported for each scenario, the measures are independent of condition and should not affect the validity of the comparisons with the NNI. The NNI scores were calculated using the smallest rectangle and convex hull methods and a Pearson's correlation was conducted for comparison.

2.2.5 Dependent Measures

The dependent measures for this study were: percent fixation for an AOI, workload based on the NNI scores, perceived workload based on the NASA-TLX, and perceived attentional control based on the attentional control survey.

3 Results

Figure 4, 5 and 6 show the % fixations calculated the AOI, main view, bird's eye view, and secondary task display respectively. The bars represent the different task loads: robotic (R), robotic and auditory monitoring (R+A), robotic and visual monitoring (R+V), and all three (R+V+A). The results show a different fixation pattern based on level of automation, AOI, and task load. For example, those in teleoperated

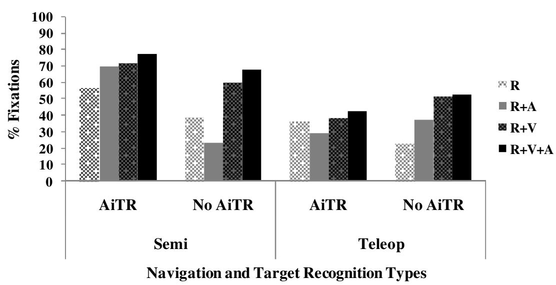


Fig. 4. % Fixations for the main viewing area

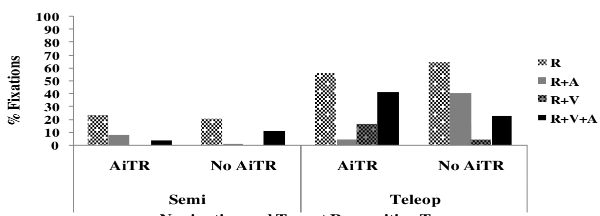


Fig. 5. % Fixations for the bird's eye viewing area

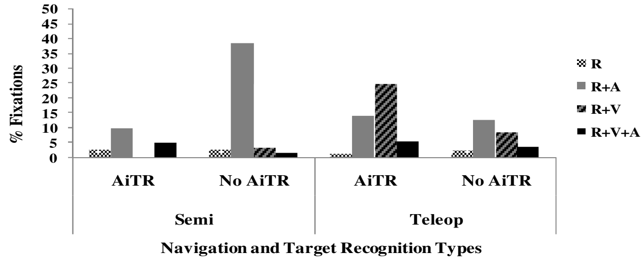


Fig. 6. % Fixations for the secondary task display for auditory and visual monitoring tasks

mode spent more time looking at the bird's eye view to navigate than looking at the main view. The participant in the semi R+A condition with no AiTR monitored the secondary display window more than the main view (Figure 6).

Participants' NASA-TLX scores are presented in Table 1. High TLX scores indicate high work load for that task. There was a significant positive correlation between the smallest rectangle and convex hull NNI calculations, r(46) = .94, p = .01. In Table 2

Table 1. Perceived work load as measured by the NASA-TLX scores

	Semi Auto		Teleop	
Task Load	AiTR	No AiTR	AiTR	No AiTR
R	40.33	39.25	44.33	35
R + V	43	25	24	39.67
R + A	5	77.67	78.67	53.33
R + V + A	52	58	72.33	74.33

Table 2. NNI calculated using the smallest rectangle method

	Semi Auto		Teleop	
	AiTR	No AiTR	AiTR	No AiTR
Task Load		Main Video		
R	0.51	0.45	0.47	0.46
R + V	0.44	0.47	0.48	0.46
R + A	0.5	0.45	0.44	0.47
R + V + A	0.46	0.47	0.5	0.48
Task Load		Birds-Eye-View		
R	0.47	0.48	0.42	0.45
R + V	0.44	0.48	0.38	0.47
R + A	0.47	0.44	0.44	0.5
R + V + A	0.46	0.48	0.48	0.47
Task Load		Secondary Task		
R	0.42	0.37	0.42	0.34
R + V	0.34	0.41	0.54	0.38
R + A	0.42	0.34	0.34	0.41
R + V + A	0.38	0.38	0.49	0.39

we present the NNI scores based on the smallest rectangle calculation since either measure provides similar comparisons. In comparing Table 2 NNI scores to Table 1 TLX scores, the overall trends match. For example, the secondary task for the R+A condition shows lower NNI scores, which match the higher perceived workload as measured by the NASA-TLX scores. However, this mapping is not perfect.

The participant in the R+V+A teleop, AiTR condition reported a high workload based on the NASA-TLX score. The NNI for the each AOI does not suggest that the participant is in high workload. The PAC score for this participant was the lowest of the group, (PAC score = 46). In contrast, the NNI for the participant in the same task load and navigation automation, but No AiTR matched their high NASA-TLX score. This participant reported a higher PAC score, (PAC score = 64).

4 Discussion

While preliminary, the results suggest that the NNI may predict workload levels as compared to the NASA-TLX. However, perceived attentional control of the operator may affect the fixation pattern such that it appears more random due to lack of attentional focus. Further, the high correlation between methods of calculating NNI suggests that it may be a robust measure. Taken together these result suggest that NNI as a means to provide adaptive automation warrants further exploration.

References

1. Barnes, M.J., Cosenzo, K.A.: Human robot teams as soldier augmentation in future battlefields: An overview. In: Schmorrow, D. (ed.) Foundations of Augmented Cognition, pp. 700–709. Lawrence Erlbaum, Mawah (2005)
2. Committee on Army Unmanned Ground Vehicle Technology, National Research Council: Technology Development for Army Unmanned Ground Vehicles. The National Academies Press, Washington, D.C. (2002)
3. Kortenkamp, D., Burridge, R., Bonasso, P., Schrenkenghost, D., Hudson, M.: An intelligent software architecture for semiautonomous robot control. In: Autonomy Control Software Workshop, Autonomous Agents 1999, Seattle, WA (1999)
4. Scerbo, M.: Adaptive automation. In: Neuroergonomics: The Brain at Work, pp. 239–252. Oxford University Press, NY (2007)
5. Scerbo, M.W.: Theoretical perspectives in adaptive automation. In: Parasuraman, R., Mouloua, M. (eds.) Automation and human performance: Theory and applications, pp. 37–63. Lawrence Erlbaum Associates, Mahwah (1996)
6. Parasuraman, R., Hancock, P.A.: Adaptive workload and control. In: Hancock, P.A., Desmond, P.A. (eds.) Workload and Fatigue, pp. 305–320. Lawrence Erlbaum Associates, Mahwah (2001)
7. Kaber, D.B., Endsley, M.R.: The effects of level of automation and adaptation on human performance, situation awareness and workload in a dynamic control task. Theoretical Issues in Ergonomics Science 5(2), 113–153 (2004)
8. Byrne, E.V., Parasuraman, R.: Psychophysiology and adaptive automation. Biological Psychology 42, 268–279 (1996)
9. Chen, J.Y.C., Haas, E.C., Pillalamarri, K., Jacobson, C.N.: Human-robot interface: Issues in operator performance, interface design, and technologies. In: ARL Technical Report, ARL-TR-3834. US Army Research Laboratory, Aberdeen Proving Ground, MD (2006)

10. Chen, J.Y.C., Terrence, P.I.: Effects of imperfect automation on concurrent performance of Gunner's and Robotics Operator's tasks in a simulated mounted environment. In: ARL Technical Report ARL-TR-4455. US Army Research Laboratory, Aberdeen Proving Ground, MD (2008)
11. Parasuraman, R., Barnes, M., Cosenzo, K.: Adaptive automation for human-robot teaming in future command and control systems. The International C2 Journal 1(2), 43–68 (2007)
12. Chen, J.Y.C., Drexler, J.M., Sciarini, L.W., Cosenzo, K.A., Barnes, M.J., Nicholson, D.: Operator workload and heart-rate variability during a simulated reconnaissance mission with an unmanned ground vehicle. In: Proceedings of the 2008 Army Science Conference (in press, 2008)
13. Land, M.F.: Fixation strategies during active behavior: A brief history. In: van Gompel, R.P.G., Fischer, M.H., Murray, W.S., Hill, R.L. (eds.) Eye movements: A window on mind and brain, pp. 76–95. Elsevier, Oxford (2007)
14. Di Nocera, F., Camilli, M., Terenzi, M.: A random glance at the flight deck: Pilots' scanning strategies and the real-time assessment of mental workload. Journal of Cognitive Engineering and Decision Making 1, 271–285 (2007)
15. Barber, D., Davis, L., Nicholson, D., Finkelstein, N., Chen, J.Y.C.: The mixed initiative experimental (MIX) testbed for human robot interactions with varied levels of automation. In: Proceedings of the 2008 Army Science Conference (in press, 2008)
16. Martinez-conde, S., Mcknik, S.L., Hubel, D.H.: The role of fixational eye movements in visual perception. Nature Reviews 5, 229–240 (2004)
17. Rayner, K.: Eye movements in reading and information processing: 20 years of research. Psychological Bulletin 124(3), 372–422 (1998)
18. Salvucci, D.D., Goldburg, J.H.: Identifying fixations and saccades in eye-tracking protocols. In: Proceedings of the Eye Tracking Research and Application Symposium, pp. 71–78. ACM Press, New York (2000)
19. Di Nocera, F., Terenzi, M., Camilli, M.: Another look at scanpath: Distance to nearest neighbor as a measure of mental workload. In: de Ward, D., Broohuis, K.A., Toffetti, A. (eds.) Developments in human factors in transportation, design, and evaluation, pp. 1–9. Shaker Publishing Maastricht, Netherlands (2006)
20. Derryberry, D., Reed, M.: Anxiety-related attentional biases and their regulation by attentional control. Journal of Abnormal Psychology 111, 225–236 (2002)
21. Hart, S., Staveland, L.: Development of NASA TLX (Task Load Index): results of empirical and theoretical research. In: Hancock, P., Meshkati, N. (eds.) Human mental workload, pp. 139–183. Elsevier, Amsterdam (1988)

Combining Electroencephalograph and Functional Near Infrared Spectroscopy to Explore Users' Mental Workload

Leanne M. Hirshfield[1], Krysta Chauncey[1], Rebecca Gulotta[1], Audrey Girouard[1], Erin T. Solovey[1], Robert J.K. Jacob[1], Angelo Sassaroli[2], and Sergio Fantini[2]

[1] Human Computer Interaction Lab
[2] Biomedical Engineering Department
Tufts University, Medford, MA 02155, USA
{Leanne.Hirshfield,Krysta.Chauncey,Rebecca.Gulotta, Audrey.Girouard,Erin.Solovey,Robert.Jacob,Angelo.Sassaroli, Sergio.Fantini}@tufts.edu

Abstract. We discuss the physiological metrics that can be measured with electroencephalography (EEG) and functional near infrared spectroscopy (fNIRs). We address the functional and practical limitations of each device, and technical issues to be mindful of when combining the devices. We also present machine learning methods that can be used on concurrent recordings of EEG and fNIRs data. We discuss an experiment that combines fNIRs and EEG to measure a range of user states that are of interest in HCI. While our fNIRS machine learning results showed promise for the measurement of workload states in HCI, our EEG results indicate that more research must be done in order to combine these two devices in practice.

Keywords: fNIRs, EEG, near infrared spectroscopy, workload.

1 Introduction

Current research in human computer interaction (HCI) explores the measurement of computer users' brain activity in an attempt to increase the effectiveness of usability testing and to create adaptive user interfaces. By measuring users' mental states objectively, and in real time, usability experts use information about users' mental workload (WL) as an additional metric during usability studies. Designers of adaptive systems hope to use this information as passive input, providing the system with reliable, real time information about the user's state, so that the system can adapt and make the human-computer interaction as flawless as possible.

Electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRs) are popular, non-invasive, brain imaging techniques. Unlike other brain devices which require subjects to lie in restricted positions (fMRI), or to drink hazardous materials (PET), EEG and fNIRs can non-invasively measure users' brain activity in real working conditions [1]. This makes EEG and fNIRs appropriate choices for brain measurement in HCI. The majority of brain measurement research in HCI uses EEG to

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 239–247, 2009.

measure users' states while a smaller body of research uses fNIRs. Only a handful of researchers have explored the combination of the two devices to measure brain function [2-5]. This is unfortunate, as combining the two devices can provide complementary information about different physiological responses to brain function and compensate for practical and functional limitations of each measurement technology.

EEG and fNIRS measure different physiological responses to mental state changes. EEG measures the electrical potentials caused by neurons firing during brain activity. fNIRs measures blood volume and oxygenation changes, reflecting hemodynamic responses to brain activity. Therefore, combining EEG and fNIRS can provide information about both the neural activations and the subsequent oxygenation and blood flow changes in the brain. Additionally, each device has functional and practical limitations that make it difficult to acquire a range of user states in real world settings. We discuss these benefits and pitfalls in detail in the next section, and we show how combining the devices allows us to compensate for the measurement pitfalls that either device has on its own.

This paper has three primary contributions to the HCI realm. First, the paper is intended as a useful guide for researchers interested in combining fNIRs and EEG. We discuss the physiological metrics that EEG and fNIRs can measure, the functional and practical limitations of each device, and technical issues to be mindful of when combining the devices. Second, we present simple machine learning methods that can be used on concurrent recordings of EEG and fNIRs data. Third, we describe results from an experiment using concurrent recordings of EEG and fNIRs.

The rest of this paper proceeds as follows: First, we discuss background literature about EEG and fNIRs. Next we present an experiment designed to elicit varying levels of working memory load on subjects. We then describe the machine learning techniques that we implemented to analyze the experiment data. After describing the results of our data analysis, we discuss our concluding thoughts and extensions for future work.

2 Background and Relevant Literature

2.1 Electroencephalography

EEG is the most studied non-invasive brain imaging device due to its fine temporal resolution, ease of use, and low set-up cost. EEG uses electrodes placed on the scalp to measure and record the electrical potential caused by neurons firing in the brain during brain activity. These measurements vary predictably in response to changing levels of cognitive stimuli [6]. Additionally, EEG benefits from high temporal resolution, enabling it to measure changes in cognitive activity on the millisecond scale. Therefore, EEG measurements are continuously reflective of a participant's cognitive states [7]. However, there are some significant limitations to EEG signals. EEG has low spatial resolution (about 10 cm) [7] , making it difficult to make precise measurements about the area of the brain being activated. EEG is also susceptible to motion artifacts, such as blinking and movement. These actions create artifacts or noise in the data which is, in some cases, stronger than the signal from the neural activity [7]. Noise is also introduced into the EEG signal from electrical interference and the subjects' breathing and heartbeat. Despite these limitations, EEG is a promising tool for the continuous measurement of cognitive states. Lee and Tan used a simple and inexpensive EEG system (~$1,500) to differentiate between various users' states that have relevance within the HCI domain [8]. They helped to bridge the gap between brain imaging research and the field of HCI, by providing detailed information about the nature of the EEG signal, its potential within the HCI field, preprocessing methods, and machine learning techniques for the EEG signal [8]. In a later paper, Grimes et al provided an overview of the pragmatics and practicality of using EEG for classification of users' working memory load [9]. They classified two working memory (WM) states with up to 99% accuracy, and four WM states with up to 88% accuracy. They discuss possible applications of EEG memory load measurements as additional metrics for usability testing, and as an additional input to adaptive systems [9]. EEG has also been used in more realistic experimental settings that apply to the military domain. For example, 'smart' EEG helmets have been designed to monitor pilots' mental state while in the air [10]. Other experiments have used portable EEG systems that monitor soldiers' mental WL while completing realistic training scenarios [11].

2.2 Functional near Infrared Spectroscopy

fNIRs has been introduced in the last two decades [12], and it is primarily used in the medical domain, and in research labs where the focus is to validate and re-design the device itself. fNIRs uses optical fibers placed on the scalp or forehead to send light in the wavelength range of 650-850 nm into the head. A small percentage of this light migrates through the scalp, skull and brain cortex and eventually is collected by other optical fibers placed 2-3 cm away from the source fibers. In the near-infrared range the main tissue absorbers are oxy-and deoxy-hemoglobin, therefore any change in the concentration of these two chromophores (as during brain function) is reflected into intensity changes at the detector's sites[12]. The spatial resolution in fNIRs is limited to approximately 5 mm. Researchers have shown that by placing the probes on a subject's forehead, fNIRs provides an accurate measure of activity within the frontal lobe of the brain, which is responsible for many high order cognitive functions, such as memory and problem solving that make up mental WL [1]. More specifically, there is a positive correlation between the increase of oxygenated blood and the increase in cognitive WL [1, 13]. These results are promising since fNIRs is portable, safe, less invasive than other imaging techniques, and has been implemented wirelessly, enabling use in real world settings [1]. However, fNIRs is not without its own limitations. Unlike PET or fMRI, fNIRs can not measure deep brain structures and it is primarily placed on the forehead, as hair can introduce noise into the signal. fNIRs is also limited by its low temporal resolution, as it take several seconds to monitor blood in the brain. Since most research in fNIRs concerns validating the tool itself, the extensive applications conducted with brain imaging techniques such as EEG, have yet to be implemented with fNIRs. Only a handful of researchers have paired the fNIRs data with machine learning techniques [14-17].

2.3 Combining EEG and fNIRs

A few researchers have explored concurrent recordings of EEG and fNIRS [2-5]. These researchers note the functional limitations of each device. By pairing the low spatial and high temporal resolution of EEG with the high spatial and low temporal resolution of fNIRs, it may be possible to overcome limitations of each measurement technology. Not only do the two devices complement each other by improving on one another's measurement pitfalls, but they also measure different physiological markers in the brain, providing further information about a user's mental state than either device could achieve alone. EEG and fNIRs data were concurrently recorded in the DARPA Augmented Cognition Technical Integration Experiment, where participants were instructed to play Warship Commander, a task which involves monitoring a radar screen for airplanes and then responding to their presence [2]. Researchers collected fNIRs and EEG data, in addition to several other types of physical and mental measurements in an attempt to compare sensor technologies [2]. Another instance in which fNIRS and EEG were used concurrently is a study conducted by Salvatori et al [5]. Due to the novel nature of this research, Salvatori's study was primarily interested in the logistics of combining the two devices and emphasis was placed on issues such as data synchronization and sensor placement. Participants in the study were asked to watch a computer screen that displayed an alternating white and black checkerboard pattern[5]. Savran et al took concurrent recordings of EEG and fNIRS data while subjects viewed images from the International Affective Picture System (IAPS) [4]. They described the creation of a database to hold EEG and fNIRs data, and they discussed issues in data synchronization and sensor placement. Researchers combining EEG and fNIRs look at the concurrent data separately, and while some initial ideas for concurrent data analysis have been discussed [4], research to date has not taken the next step of actually combining EEG and fNIRS data during data analysis. To the best of our knowledge, we are the first to attempt classification of concurrent recordings of EEG and fNIRs data with machine learning.

3 Experiment

3.1 Equipment

The fNIRs device used in this study is an OxyplexTS (ISS Inc. Champagne, IL) frequency-domain tissue spectrometer with two optical probes. Each probe has a detector and four light sources (Fig. 1c). Each light source emits near infrared light at two separate wavelengths (690nm and 830nm) which are pulsed intermittently in time. This results in 2 probes x 4 light sources x 2 wavelengths = 16 light readings at each timepoint (sampled at 50Hz). Electroencephalograms were collected using 32-channel caps.

Electrodes were arranged according to the International 10-20 system (1b). The EEG was amplified using an SA Bioamplifier (SA Instruments, San Diego Ca.) with a bandpass of .01 and 40 Hz. The computer sampled at a rate of 200 Hz. The fNIRs probes were placed on subjects' forehead using an athletic headband, leaving room for the electrode cap to fit on their heads (1a).

3.2 Experiment Tasks

Our experiment had three conditions (Fig. 2). In each condition, users viewed rows of randomly generated red and blue planes moving down a screen.

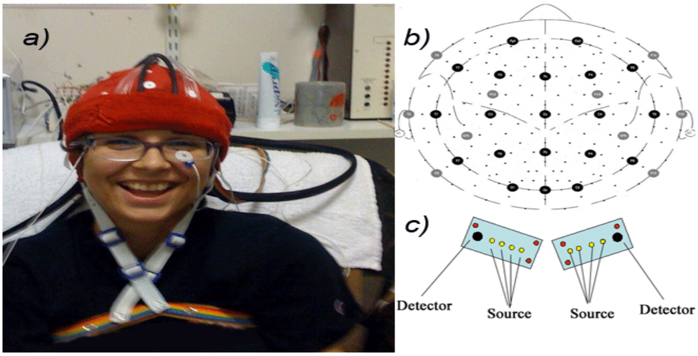


Fig. 1. a) A subject wears fNIRs/EEG setup. b) electrode placement c) fNIRs sensors.

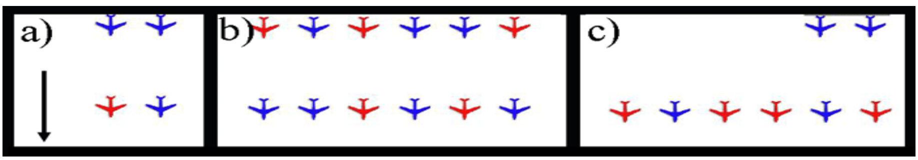


Fig. 2. The three experiment conditions are low WL (a), high WL (b) and random WL (c)

Users kept track of the total number of red and blue planes that they had seen, and after 60 seconds, they were prompted to give the total count. While the figure below shows 2 rows of planes in each condition, in reality, only one row of planes was on the screen at any given time. In the first condition (Figure 2a) subjects made WM updates based on the two planes that were in each row. In the second condition (2b) subjects viewed six planes per row, and in the final condition (2c), subjects viewed randomly generated rows of two or six planes. We based our WL conditions on the fact that set size has been used to manipulate WM for decades. We will refer to these conditions as the low WL (2 planes per row), random WL (2 or 6 planes per row), and high WL (6 planes per row) conditions.

3.3 Task Events and Synchronization

Events were automatically generated and logged each time that one of the 60 second long tasks began or ended. Ideally, the EEG and fNIRs data acquisition would be synchronized in time, and any events recorded throughout the experiment would be logged in the concurrently recorded EEG and fNIRs acquisition systems [4, 5]. Due to hardware limitations, an approximation was made. The EEG and fNIRs acquisition systems were set to start at the same time via two button presses (one for the fNIRs acquisition system and one for the EEG system). Each event was logged directly into the fNIRs acquisition system, which sampled at 50Hz, and added to the EEG data after the experiment ended.

3.4 Experiment Methodology

Four right handed, undergraduate students at Tufts completed this experiment (3 women, 1 male). Subjects were instructed to keep movement to a minimum, and to count the number of red and blue planes that they saw during each 60 second task period. After each task ended, they verbally gave their answers. Subjects rested for 20 seconds between each task. We used a randomized block design with 8 trials, resulting in 24 tasks.

4 Results and Analysis

4.1 Performance Results

We calculated subjects' performance (whether they said the correct number of red/blue planes) on the three experiment conditions. WM load increased as the number of planes per row increased, and subjects' performance on each of the experiment conditions reflected this increase in WL. Subjects had the highest accuracy for the tasks that involved 2 planes per row (low WL). They had the lowest accuracy on tasks with 6 planes per row (high WL), and their accuracy on the random WL condition was always between the accuracy on the high and low WL conditions.

4.2 Data Preprocessing and Machine Learning Analysis

As brain activity differs widely on a person to person basis, we run all analyses separately for each subject. We developed machine learning techniques to classify the EEG and fNIRs data from the experiment. *Our analysis is not intended to compare the two techniques of brain measurement or the machine learning techniques.*

fNIRs: Each experiment lasted about 35 minutes, with fNIRs data recorded every .02 seconds. We recorded 16 channel readings at each timepoint, where we refer to the readings of one source detector pair at one wavelength, as one *channel*. We first normalize the intensity data in each channel by their own baseline values. We then apply a moving average band pass filter to each channel (saving frequency values between .1 and .01 Hz) and we use the modified Beer-Lambert Law[12] to convert our light intensity data to measures of the change in oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (Hb) concentrations in the brain. This results in eight readings of HbO and eight readings of Hb data at each timepoint in the experiment. To choose the best HbO and Hb channels for classification we use the techniques described in [14].We implemented a weighted k-nearest-neighbor classifier (k=3) with a distance metric computed with Symbolic Aggregate Approximation (SAX). SAX creates a symbolic approximation of time series data, allowing for dimensionality reduction. For a review of SAX, see [18].

EEG: We implemented signal processing and feature generation schemes that have been used with EEG data previously [8]. We split the continuous EEG data into small overlapping windows and we took a Fourier transform of the data in each window. We chose a window size of 2 seconds, with windows overlapping every second. For each window, we compute the magnitude and phase of the signal and the spectral power of the signal in the delta (1-4Hz), theta (4-8Hz), alpha (8-12Hz), beta-low (12-20Hz), beta-high (20-30Hz), and gamma (30-50Hz) frequency bands. We also compute the coherence and cross spectrum between each channel for each frequency band in each window. This results in over 6,200 features for each instance. We use blocked cross validation to select our most relevant attributes for classification. We use an information gain heuristic followed by Weka's *CfsSubsetEval* function to choose the features that best predict the class label in the training data. We then apply a Naïve Bayes classifier to data.

4.3 Classifying Working Memory Load

We attempted to classify each 60 second long period of time when subjects were completing one of the 3 conditions described in Fig 2. We classified these conditions with our fNIRs data, and with our EEG data. Classification results are depicted in Table 1. We looked at our ability to classify low and high WL, low and random WL, random and high WL, and low, random, and high WL. When using just the fNIRs data for analysis, we see that we achieved promising accuracy for each subject. However, with the exception of subject 1, the EEG data yields nearly random accuracy. It is promising that the fNIRs classification yielded accuracy as high as 82% distinguishing two WL classes and up to 50% for distinguishing three WL classes.

Table 1. Classification accuracy on just fNIRs data (unshaded) and just EEG data (shaded columns) for the three conditions in Fig 4. S1 = subject 1, etc

	random v. low	random v. low	random v .high	random v. high	low v. high	low v. high	low v. random v. high	low v random v. high
s1	57%	63%	64%	75%	61%	69%	45%	42%
s2	53%	56%	53%	69%	51%	69%	34%	46%
s3	50%	82%	52%	69%	49%	75%	33%	50%
s4	52%	78%	50%	65%	52%	75%	34%	46%

4.4 Analysis of Results

There are many possible explanations for the low EEG classification accuracy in this experiment. Our ability to accurately measure users' states depends on a number of factors. Any changes in the placement of the fNIRs and EEG sensors on subjects, changes in the analysis techniques applied to the brain data, or changes in the experiment tasks could result in higher or lower classification accuracy of fNIRs or EEG brain data. It is possible that the chosen task elicited brain activity primarily in the prefrontal cortex, which was dominated by the fNIRs sensors. It is also possible that the fNIRs light sources introduced noise into the EEG signal that was difficult to remove [4]. Therefore, the EEG sensors may not have been able to pick up the resulting brain activity.

5 Conclusion and Future Work

In this paper, we discussed the physiological metrics that EEG and fNIRs can measure, the functional and practical limitations of each device, and technical issues to be mindful of when combining the devices. We presented machine learning methods that can be used on EEG and fNIRs data. We presented an experiment designed to combine fNIRs and EEG to measure three user states. While the EEG results were low, we demonstrated the ability of fNIRs to classify mental WL states. fNIRs is a relatively new device, and it holds great potential for the HCI domain. Future work will explore factors that may have contributed to the low EEG results. It is possible that EEG and fNIRS can best complement eachother when subjects complete tasks that activate more than just the prefrontal cortex. It is apparent that EEG and fNIRs, when combined, have the potential to acquire complementary information about user states. However, more research is needed to achieve this goal.

References

1. Izzetoglu, K., et al.: Functional Optical Brain Imaging Using Near-Infrared During Cognitive Tasks. International Journal of Human-Computer Interaction 17(2), 211–231 (2004)
2. John, M.S., et al.: Overview of the DARPA Augmented Cognition Technical Integration Experiment. International Journal of Human-Computer Interaction 17(2), 131–149 (2004)
3. Tong, Y., et al.: Concurrent recordings of electrical evoked potentials and near-infrared responses to brain activation. Tufts University (2007)
4. Savran, A., et al.: Emotion Detection in the Loop from Brain Signals and Facial Images. In: eNTERFACE 2006, Dubrovnik, Croatia (2006)
5. Salvatori, G., et al.: Combining Near-Infrared Spectroscopy and Electroencephalography to Monitor Brain Function. In: Instrumentation and Measurement Technology Conference. IEEE, Sorrento (2006)
6. Anderson, C., Sijercic, Z.: Classification of EEG Signals from Four Subjects During Five Mental Tasks. In: Solving Engineering Problems with Neural Networks: Proceedings of the Conference on Engineering Applications in Neural Networks, pp. 407–414 (1996)
7. Berka, C., et al.: Real-Time Analysis of EEG Indexes of Alertness, Cognition, and Memory Acquired With a Wireless EEG Headset. International Journal of Human Computer Interaction 17(2), 151–170 (2004)
8. Lee, J.C., Tan, D.S.: Using a Low-Cost Electroencephalograph for Task Classification in HCI Research. In: ACM Symposium on User Interface Software and Technology (2006)
9. Grimes, D., et al.: Feasibility and Pragmatics of Classifying Working Memory Load with an Electroencephalograph. In: CHI 2008 Conference on Human Factors in Computing Systems, Florence, Italy (2008)
10. Gevins, A., et al.: Flight helmet EEG system, in Final Tech Report AL/CF-SR-1993-0007. Sam Technology, San Fracisco, CA (1993)
11. Mathan, S., et al.: Neurophysiological Estimation of Interruptibility: Demonstrating Feasibility in a Field Context. In: Proceedings of the 4th International Conference of the Augmented Cognition Society, Baltimore, MD (2007)
12. Chance, B., et al.: A novel method for fast imaging of brain function, non-invasively, with light. Optics Express 10(2), 411–423 (1988)
13. Parasuraman, R., Caggiano, D.: Neural and Genetic Assays of Human Mental Workload. In: Quantifying Human Information Processing, Lexington Books (2005)

14. Hirshfield, L.M., et al.: Brain Measurement for Usability Testing and Adaptive Interfaces: An Example of Uncovering Syntactic Workload in the Brain Using Functional Near Infrared Spectroscopy. In: Conference on Human Factors in Computing Systems: Proceeding of the twenty-seventh annual SIGCHI conference on Human factors in computing systems (2009)
15. Sassaroli, A., et al.: Discrimination of mental workload levels in human subjects with functional near-infrared spectroscopy. The Journal of Innovative Optical Health Sciences (accepted, 2009)
16. Son, I.-Y., et al.: Human performance assessment using fNIR. In: Proceedings of SPIE The International Society for Optical Engineering, vol. 5797, pp. 158–169 (2005)
17. Sitaram, R., et al.: Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain-computer interface. NeuroImage 34(1416) (2007)
18. Lin, J., et al.: A Symbolic Representation of Time Series, with Implications for Streaming Algorithms. In: Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, San Diego, CA (2003)

Detecting Intentional Errors Using the Pressures Applied to a Computer Mouse

Curtis Ikehara and Martha E. Crosby

University of Hawaii at Manoa, 1680 East-West Road, Honolulu, HI 96822, USA
{Crosby,Cikehara}@hawaii.edu

Abstract. Intentional errors are considered a form of deceit. In this pilot study, the pressures applied to a computer mouse will be analyzed to determine if it is possible to detect intentional errors. Twenty participants ranging in age from 18 to 21 years performed a task involving intentionally making errors when instructed. A comparison will be made between the pressures applied to a computer mouse when answering the questions with the intention of being correct and with the intention of making an error. The data will need to be normalized for each individual to obtain accurate results. The analysis of the pressures may indicate that there are detectable variations within some individuals. Due to the preliminary nature of this study further research will be required.

Keywords: Intentional errors, deceit, pressure sensitive computer mouse.

1 Introduction

Deception is a broad category of acts to convince another to believe false information. Within the category of deceitful acts, a lie can be a statement one knows as false with the intention of the statement being taken as the truth. Studies using functional magnetic resonance imaging show that different areas in the brain are active during deceptive acts [1, 2]. Evidence that mental acts of deception can influence physiological factors include a variety of physiological measures used to detect deception (e.g., polygraph – breathing, electro-dermal activity and cardiovascular activity), but other emotions unrelated to lying can cause physiological responses that confound the detection of deception [3]. Facial expressions, derived from the facial muscles, have been used to detect deceit [4]. The pressures, derived from finger and hand muscles, applied to a pressure sensitive computer mouse has been used to evaluate the mental activity of a user's cognitive load [5, 6]. In this pilot study, a comparison will be made between the pressures applied to a computer mouse when answering questions with the intention of being correct and with the intention of making an error.

For this study, an intentional error is when a person purposefully provides the wrong answer when instructed. Being instructed to make an error reduces the emotion associated with trying to deceive another person, but the intent of making an error, or lying, is still present.

Walczyk, Roper, Seemann and Humphre [8] posited three cognitive stages of lying which are activation (i.e., question is received and truth accessed), decision to lie, and

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 248–253, 2009.

construction of lie. The three stages typically are in sequence. For this pilot study, questions are being presented on a computer-based multiple choice response system that instructs the person when to lie and provides the person with false answers fixing the time spent on these stages.

A person can approach making an intentional error in several ways. An example would be after the person determines the correct answer in the activation stage and following the instruction to lie in the decision stage the person would select the first wrong answer, but this may not necessarily be the order of mental activity. The person may first follow the lie instruction in the decision stage, which could then interfere with the activation stage (i.e., determining the correct answer) making the task of finding a solution more cognitively difficult and possibly increasing the time spent in the activation stage.

A comparison will be made between the pressure patterns applied to a computer mouse when answering the questions with the intention of being correct and with the intention of making an error. It is hypothesized that the instruction to give the wrong answer (i.e., to lie) can for some people increase their cognitive load which can be detected in the patterns of pressures applied to a pressure sensitive computer mouse.

2 Method

Twenty participants performed a task involving intentionally making errors when instructed on a computer system. They used a pressure sensitive computer mouse (PSCM) to answer the questions. Data from the PSCM was recorded along with the participant responses.

2.1 Participants

Twenty volunteers, 18 males and 2 females, from the US Air Force Academy ranging in age from 18 to 21 years, participated in the study. Each volunteer was tested individually during an approximately one hour session where the task for this experiment was conducted after two previous unrelated experimental tasks.

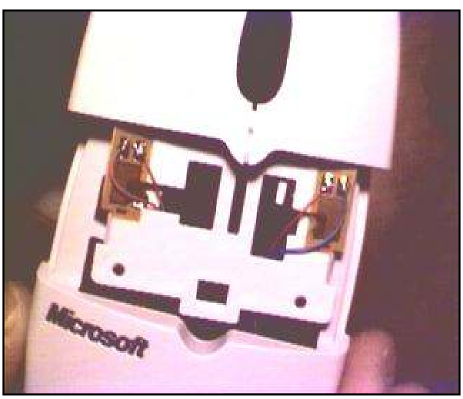


Fig. 1. Computer mouse with pressure sensors on buttons

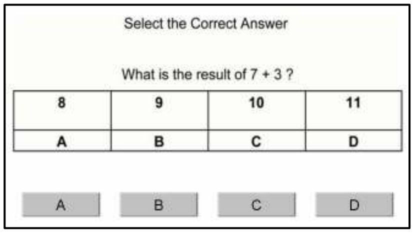


Fig. 2. A sample view of a question presented to the participant. Buttons to select the correct answer are at the bottom of the screen.

Table 1. An example of one of the three sets of questions

	Instruction	Question	A	B	C	D
1	Select the Correct Answer	What is the result of 7 + 3?	8	9	10	11
2	Select the Correct Answer	What word that has the same meaning as "jump."	bean	leap	rope	tool
3	Select the Correct Answer	What is the result of 7 + 17 + 3 + 5?	31	32	33	34
4	Select your Best Guess	What is the result of 1472 times 27?	20000 to 30000	30000 to 40000	40000 to 50000	50000 to 60000
5	Select your Best Guess	What word has the same meaning as "hold"?	grip	clasp	keep	contain
6	Select the Wrong Answer	What is the result of 7 + 3?	8	9	10	11
7	Select the Wrong Answer	What word that has the same meaning as "jump"?	bean	leap	rope	tool
8	Select the Wrong Answer	What is the result of 7 + 17 + 4 + 5?	31	32	33	34

2.2 Equipment

A custom designed computer mouse equipped with pressure sensors inside the body and the buttons of the mouse was used to detect the pressures applied during task performance. Data of the pressures applied to the computer mouse while clicking on answers during the task was collected (see Figure 1).

2.3 Task

The task involved answering three sets of questions (see Table 1). Each set consisted of eight questions with four possible responses. The first three questions were simple math or language questions. The next two questions are difficult or ambiguous. Note that for question 5 in Table 1, all the answers mean hold in different context. The last three were similar to the first three, but the participant was instructed to give a wrong answer. Each question was displayed individually on the screen (see Figure 2).

3 Results

Analysis: A comparison will be made between the pressures applied to a pressure sensitive computer mouse (PSCM) when answering the questions with the intention

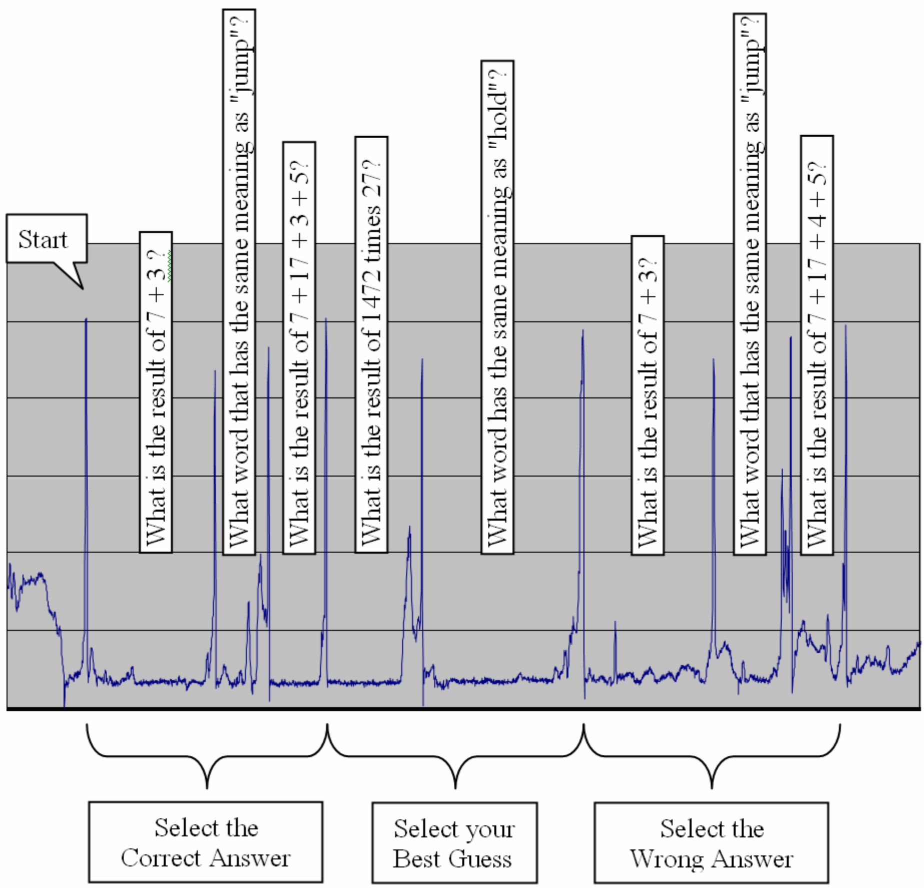


Fig. 3. Shown are the pressures applied to the computer mouse, the instructions to the participant and the questions being asked

of being correct and with the intention of making an error (see Figure 3). Previous pilot studies indicated that minimal difficulty or cognitive load is indicated by a sudden sharp pressure on the mouse button when selecting a response while high difficulty is indicated by a distortion in the shape of the pressure on the mouse button. It is hypothesized there will be a differential between the pressure patterns applied to a PSCM when the intention is to be correct versus the intention of making an error. To obtain the optimal differentiation of patterns, the pressures applied to the PSCM will need to be normalize to the pressure characteristics of each individual. Previous studies have found pressure variations unique to the individual [9, 10].

4 Discussion

Analysis of the pressures on a computer mouse will indicate that there may be detectable variations within some individuals and it is likely that the data needs to be normalized for each individual to optimize categorization of the pressure patterns when a person is giving the correct answer and giving the wrong answer. Should the results of this study support the hypothesis, the pressures applied to a computer mouse could be used for detecting deception.

When detecting deception, an interviewer's personality can affect the response of a person being questioned making it difficult to determine if deception is occurring. Since it is common for people to answer questions on a computer using a mouse it is possible to minimize the variable emotional impact of an interviewer on the interviewee. A pressure sensitive computer mouse (PSCM) that via pressure patterns and cognitive load can detect deceit would be a valuable asset for security screening. Due to the preliminary nature of this study further research will be required.

Acknowledgments. We would like to thank the United States Air Force Academy and especially Maj. Heather Pringle and Maj. Carlene Perry for their extraordinary support with the experiment. This research was supported in part by the Office of Naval Research grant no. N00014970578 and DARPA grant no. NBCH1020004.

References

1. Bhatta, S., Mbwanab, J., Adeyemob, A., Sawyerb, A., Hailub, A., Van Meterb, J.: Lying about facial recognition: An fMRI study. Brain and Cognition 69(2), 382–390 (2009)
2. Gordon, N.J., Mohamed, F.B., Faro, P.S.: Integrated zone comparison polygraph technique accuracy with scoring algorithms. Physiology & Behavior 87, 251–254 (2006)
3. Ford, E.B.: Lie detection: Historical, neuropsychiatric and legal dimensions. International Journal of Law and Psychiatry 29(3), 159–177 (2006)
4. Ekman, P., Friesen, W.V., O'Sullivan, M.: Smiles when lying. Jnl. of Personality and Social Psychology 54(3), 414–420 (1988)
5. Ikehara, C.S., Crosby, M.E.: Using Real-Time Physiological Monitoring for Assessing Cognitive States. In: Ghinea, G., Chen, S. (eds.) Digital Multimedia Perception and Design, ch. 8, pp. 170–185 (2006)
6. Ikehara, C.S., Crosby, M.E., Chin, D.N.: A Suite of Physiological Sensors for Assessing Cognitive States. In: 11th International Conference on Human-Computer Interaction (July 2005)

7. Ikehara, C.S., Crosby, M.E.: Assessing Cognitive Load with Physiological Sensors. In: 38th Annual Hawaii International Conference on System Sciences (January 2005)
8. Walczyk, J.J., Roper, K.S., Seemann, E., Humphrey, A.M.: Cognitive Mechanisms Underlying Lying to Questions: Response Time as a Cue to Deception. Appl. Cognit. Psychol. 17, 755–774 (2003)
9. Crosby, M.E., Ikehara, C.S.: Biometric Technology for Human Identification Continuous Identity Authentication Using Multimodal Physiological Sensors. In: SPIE Defense and Security Symposium (April 2004)
10. Ikehara, C., Crosby, M.E.: User Identification Based on the Analysis of the Forces Applied by a User to a Computer Mouse. In: Proceedings of the Hawaii International Conference on System Sciences, Kona, Hawaii (January 2003)

Visual Navigation Patterns and Cognitive Load

Laurel A. King

Communication and Information Sciences, University of Hawaii at Manoa
1680 East-West Road, POST 317, Honolulu, HI 96822
laurel.king@hawaii.edu

Abstract. Eye tracking technology is a prospective tool for augmenting cognition in real-time in response to screen navigation and other eye movements that can be monitored. This paper examines eye movements associated with differences in problem complexity. The experiment utilized constraint satisfaction problems of differing difficulty measured by the number of steps necessary to complete and the relative time required to solve it. Participants were observed and tested through an eye-tracking experiment to see if correlations between visual navigation and problem complexity were present. Eye movement patterns, in particular pupil size, have been used to measure cognitive load in other contexts [6-9]. The results showed overall increases in fixations and pupil size that corresponded to increases in problem complexity.

Keywords: Cognitive load, eye tracking, analytical reasoning.

1 Introduction

This paper explores the consequences of different levels of complexity on screen navigation during problem-solving tasks. Participants were observed and tested through an eye-tracking experiment to determine whether correlations between visual navigation and problem complexity could be detected and measured. In addition, individual differences in domain expertise, short-term memory, spatial ability, cognitive style and learning preference, which are thought to affect internal and external problem representation and comprehension of prepared external representations of problems, were measured [1-5] for possible affect as well. Eye movement patterns, in particular pupil size, have been used to measure cognitive load in other contexts [6-9]. If differences in visual navigation and other measures such as the duration and location of fixations and changes in pupil size can be observed using eye tracking, these differences can be used to augment cognition or customize views appropriately as eye tracking and other monitoring devices improve. Changes in pupil size are easier to track with eye tracking cameras because calibration is not required.

To better understand how increases in cognitive load due to an increase in task difficulty are present in the visual process and the strategy of the individual, the following question was considered. Do different eye tracking patterns occur for different levels of complexity in analytical reasoning? This was examined within subject as well as between subjects to see if measurable trends exist that could be used to augment cognition and lessen cognitive overload for a specific user. This would be

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 254–259, 2009.

particularly important for kiosks and computers in schools and public places that are used by people of varying levels of expertise.

In order to measure changes related to increased cognitive load, eye movements were tracked as participants attempted to find a missing entry on a series of five simple 2x2 sudoku puzzles. The placement of the missing entry ranged in difficulty by the number of entries that must be remembered in order to deduce the missing entry. The goal was to see if changes in visual process corresponding to changes in complexity could be detected from eye movement patterns within subjects.

Problem complexity was defined in two basic ways in the study. Initially, problems were ranked for complexity according to the number of steps necessary to solve them. All of the problems were constraint satisfaction problems and grouped according to this definition of difficulty. The second measure of difficulty was calculated by the actual time taken to complete the problem and they were ranked accordingly. This method allowed for each participant to have his or her own unique ranking, but problems were also averaged over all participants to give an overall difficulty ranking based on duration. These rankings were used to compare visual navigation of similar problem types based on these two measures of difficulty.

Measuring and understanding differences that correlate to changes in complexity during problem solving can further our understanding of how representations can be customized to improve problem solving and recognize cognitive load for individuals and groups. The ability to identify cognitive load levels using a passive monitoring technique or sensors would benefit interface design and artificial intelligence in a variety of fields including education and business.

2 Previous Research

Cognitive overload presents problems for users as well as learners. As problem complexity increases, adequate problem representation becomes very important and often determines how successful the person will be in solving a problem [1, 2]. Jonasson states that qualitative representation of a problem prior to the quantitative solution indicates deeper conceptual understanding of the problem, whereas attempting to apply formulas without a qualitative and structural understanding of the problem is less effective and typical of novices[3]. Differences in visual navigation can indicate whether the problem solver is using a strategy that indicates structural understanding of the problem, leading to lower cognitive load from more efficient visual navigation and problem solving. Also, when problems require collaborative and interdisciplinary solutions, successful communication of a problem representation grows in importance [4, 5]. This research can also lead to better understanding of the effectiveness of representations for visual communication. If cognitive overload can be measured automatically, the information can be used to mitigate this through various techniques such as scaffolding, multiple problem representation, worked problems, etc.

Eye tracking represents a promising technique for measuring cognitive load. It is thought that fixation duration is a measure of difficulty of information extraction and interpretation; while the number of fixations in a region indicates level of interest [6]. The pupil size can also indicate things about the viewer and level of cognitive activity. Eye movements and changes in pupil size can reveal whether a person is experiencing cognitive overload [7]. Kahneman's theory of attention is partially based on the relationship between cognitive activity and pupil dilation [8]. Differences in pupillary response is seen as the most promising physiological indicator of cognitive load, although it may be less reliable for older users [9]. This methodology was further investigated using *sudoku* puzzles, which are examples of diagrammatic constraint satisfaction problems. Since they require no reading to complete, they are useful for measuring cognitive load in diagrammatic contexts, yet the semantics and rules are very simple and easy to remember.

3 Methodology

Eye movements were tracked and analyzed as seven university students completed a series of five simple *sudoku* analytical problems of varying complexity. The participants were also tested for other individual differences related to diagrammatic communication and reasoning, such as working memory capacity, visual and verbal preferences for learning and thinking, and spatial ability. This information was used to analyze possible correlations between these factors, strategies used, and indications of resulting cognitive load.

The goal of the study was to see if changes in cognitive load corresponding to problem complexity could be observed and recognized in eye movement patterns. This was analyzed quantitatively in terms of changes in pupil size and the number and duration of eye fixations. It was also analyzed qualitatively in terms of navigation patterns and the strategy used to solve problems. These were examined to understand the different solution strategies used and how the level of cognitive load impacted the strategy or process within and between subjects.

The following set of hypotheses was used to test for changes in eye-movements and pupil size correlating to differences in problem difficulty. The participants completed one practice *sudoku* problem and five *sudoku* problems of varying difficulty. Difficulty was measured in terms of the number of steps that must be remembered in order to find the correct solution. Problems increased in difficulty level according to the number of steps required to solve it. This was used to see if corresponding differences in the number of fixations, duration of fixations, average fixation, and average pupil size could be measured. These differences were also evaluated in terms of the other independent variables, such as the visual-verbal factors from the questionnaire, the spatial test and memory test scores.

Ha: Participants will have more eye fixations while solving higher difficulty problems than lower difficulty problems.

Hb: Participants will have longer duration of eye fixations while solving higher difficulty problems than lower difficulty problems.

Hc: Participants will have higher average pupil sizes while solving higher difficulty problems than lower difficulty problems.

Participants completed constraint satisfaction problems using a computer while their time per screen, fixation locations and durations, pupil size, and answers to problems were recorded using EventStream software and an ASL eye tracker. Their eye movements were tracked sixty times per second. The information from EventStream on the task completion times, answers to problems, and fixation locations, durations of fixations by area of screen, and pupil size from the eye tracker was coded by participant and kept in spreadsheets for further processing. The data included the response time, accuracy, pupil size, the number of fixations and gazes on different parts of the display for each problem. A qualitative analysis was done to understand the participant's strategy and use of the graphic representation in relation to accuracy and response time.

All problems were designed to be solvable while viewing a computer screen without additional notes or calculations, although for increased difficulty some required more working memory capacity than others.

4 Results

The eye movements showed trends for increased fixations, longer fixation durations, and increased pupil size as complexity increased. Figure 1 shows the increase in fixations as the problem difficulty increases. Despite individual differences in the number of fixations, an overall trend of the average of all subjects shows an increase as problem difficulty increased. Figure 2 highlights the increased average fixation duration as complexity increased. Although less pronounced, the tendency is for the duration of fixation to increase with problem complexity. Figure 3 shows the change in average pupil size for the problems as they increased in complexity. Pupil size is measured using a raw data coding to show the relative changes, but since each participant sat at

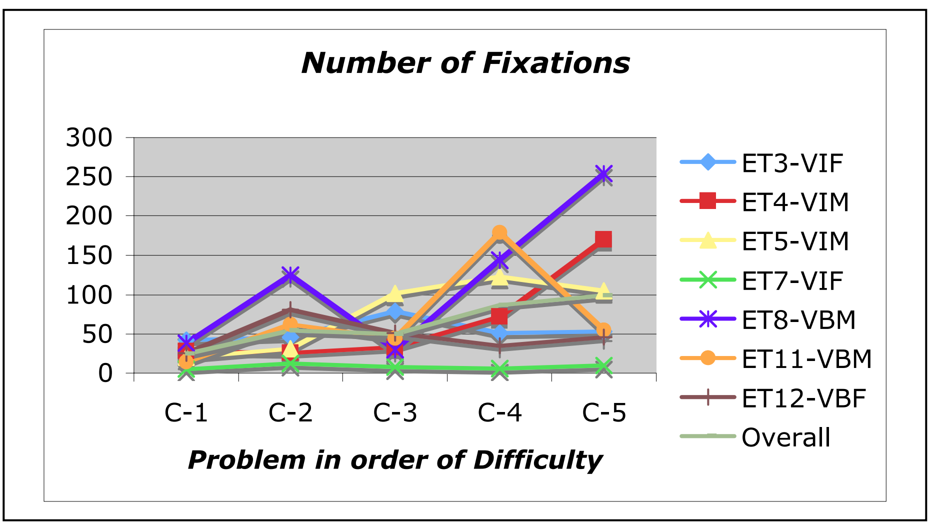


Fig. 1. Number of fixations per subject as problem difficulty increased

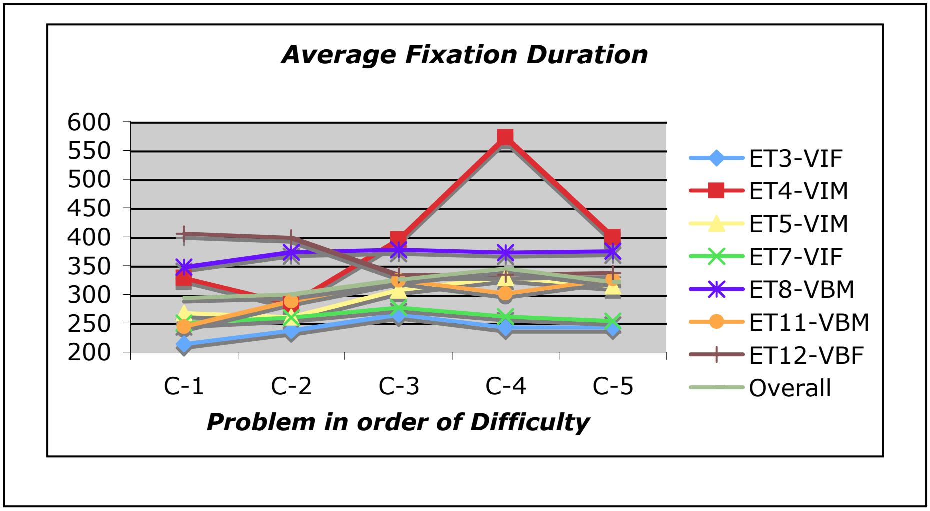


Fig. 2. Average fixation duration per subject as problem difficulty increased

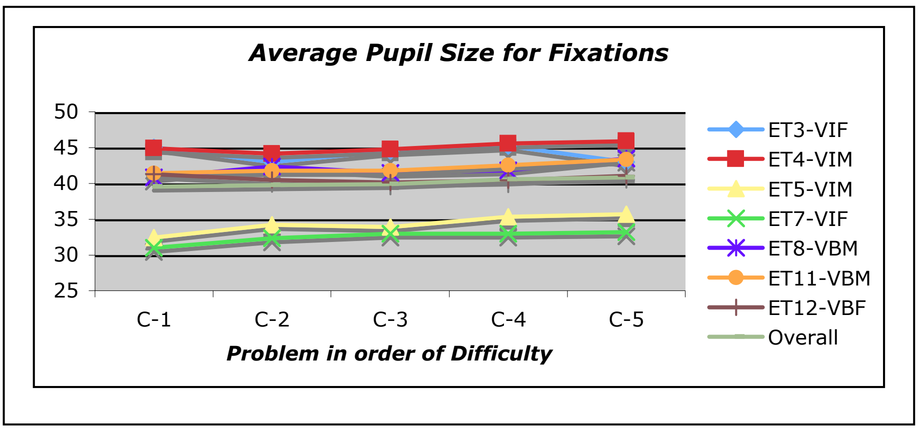


Fig. 3. Average pupil size per subject as problem difficulty increased

slightly different distance from the screen, it does not represent an exact measurement in millimeters.

Although, the change in average pupil size is small, there is a consistent upward trend for all subjects as the problem difficulty increased. The average was based on the pupil size averaged for each fixation. The blinks which caused the pupil size to be measured as zero were not used in the calculation. The wider range in pupil size seemed to be a better indicator of cognitive load than the average size.

The pupil size fluctuated throughout the problem solving tasks. The standard deviation increased as the difficulty increased as well, but the relationship was much noisier. Qualitative examination of the video of the eye movements and the pupil size seemed to indicate that the pupil size increased in connection with increased search behavior. However, this relationship is still being researched for a more specific and linkage. Intuitively, reacting to increased cognitive demand by increased search for relevant information seems plausible.

The three hypotheses were supported by the trends in the sample. The differences were not great, but there was a consistent increase in the average pupil size as well as the gap between the minimum and maximum pupil size. Eye tracking, in particular, changes in pupil size, may be feasible for detecting increases in cognitive load for augmenting cognition or determining performance level remotely.

5 Conclusion

Results showed that increases in cognitive load due to an increase in task difficulty were observed in the visual process and strategy of most participants. There is support for different eye tracking patterns to occur under increased complexity. This phenomena was examined within-subjects to see if measurable trends existed that could be used to augment cognition for a specific user during a visual task that requires reasoning. This type of user modeling and detection of cognitive load could be particularly useful for systems and interfaces that are used by people of varying levels of expertise, such as in schools and other public places.

Further research is underway to better understand how this and other visual indicators can be used to augment cognition and improve user experience with information and communication technology. Pupil size changes have potential for use in usability testing as well. Since tracking pupil size requires less calibration and is easier to do than eye movement tracking, it has potential for use with Web cameras that have become ubiquitous with Web 2.0 technologies.

References

1. Greeno, J.G.: Process of understanding in problem solving. In: Castellan, N.J., Pisoni, D.B., Potts, G.R. (eds.) Cognitive theory, vol. 2, pp. 43–83. Erlbaum, Hillsdale (1977)
2. Novick, L.R., Bassok, M.: Problem Solving, ch. 14. In: Holyoak, K.J., Morrison, R.G. (eds.) The Cambridge Handbook or Thinking and Reasoning, pp. 321–350. Cambridge University Press, Cambridge (2005)
3. Jonassen, D.: Using cognitive tools to represent problems. Journal of Research on Technology Education 35, 362–381 (2003)
4. Narayanan, N.H.: Technical Report CSE97-06: Diagrammatic Communication: A Taxonomic Overview. Auburn University (1997)
5. Van Bruggen, J.M., Kirschner, P.A.: Designing external representations to support solving wicked problems. In: Andriessen, J., Baker, M., Suthers, D. (eds.) Arguing to Learn: Confronting cognitions in computer-supported collaborative learning environments, vol. 1, pp. 177–203. Kluwer Academic Publishers, Norwell (2003)
6. Fitts, P.M., Jones, R.E., Milton, J.L.: Eye movements of aircraft pilots during instrument-landing approaches. Aeronautical Engineering Review 9, 24–29 (1950)
7. Marshall, S.P.: The Index of Cognitive Activity: Measuring cognitive workload. In: Proceedings IEEE 7th Conference on Human Factors and Power Plants, pp. 7.5–7.9. IEEE, New York (2002)
8. Kahneman, D.: Attention and effort. Prentice-Hall, Englewood Cliffs (1973)
9. Van Gerven, P.W.M., Paas, F.G.W.C., Van Merrienboer, J.J.G., Schmidt, H.G.: Cognitive load theory and aging: effects of worked examples on training efficiency. Learning and Instruction 12, 87–105 (2002)

Modeling the Cognitive Task Load and Performance of Naval Operators

Mark A. Neerincx[1,2], Stefan Kennedie[1,5], Marc Grootjen[2,3], and Franc Grootjen[4,5]

[1] TNO Human Factors, Kampweg 5, 3769 ZG Soesterberg, The Netherlands
[2] Delft University of Technology, Mekelweg 4, 2628 GA Delft, The Netherlands
[3] Defense Materiel Organization, Directorate Materiel Royal Netherlands Navy, P.O. Box 20702, 2500 ES The Hague, The Netherlands
[4] Donders Centre for Cognition, Radboud University Nijmegen, P.O. Box 9104, 6500 HE, Nijmegen, The Netherlands
[5] Radboud University Nijmegen, The Netherlands
mark.neerincx@tno.nl, Kennedie@gmail.com, marc@grootjen.nl, grootjen@acm.org

Abstract. Operators on naval ships have to act in dynamic, critical and high-demand task environments. For these environments, a cognitive task load (CTL) model has been proposed as foundation of three operator support functions: adaptive task allocation, cognitive aids and resource feedback. This paper presents the construction of such a model as a Bayesian network with probability relationships between CTL and performance. The network is trained and tested with two datasets: operator performance with an adaptive user interface in a lab-setting and operator performance on a high-tech sailing ship. The "Naïve Bayesian network" tuned out to be the best choice, providing performance estimations with 86% and 74% accuracy for respectively the lab and ship data. Overall, the resulting model nicely generalizes over the two datasets. It will be used to estimate operator performance under momentary CTL-conditions, and to set the thresholds of the load-mitigation strategies for the three support functions.

Keywords: mental load, emotion, Bayesian networks, cognitive engineering, Defense and Space operations.

1 Introduction

Crews on naval ships have to operate in dynamic, critical and complex task environments, which impose high fluctuations of the required cognitive resources. These resources are constrained and may not fit the momentary task demands, resulting in performance decrements. To mitigate such load bottlenecks, three operator support functions are being developed: adaptive task allocation, cognitive aids and resource feedback [1, 2, 3]. Important foundations of these support functions are *situated* theories on cognitive task load (CTL) and emotional state (ES) [4]. Such theories include accepted features of cognition such as limited processing capacity, are validated in the context of a specific domain and possibly group of task performers, and

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 260–269, 2009.

provide predictions of the task performance within this domain. Consequently, they can provide the "context-awareness" for the proposed support functions. Face validity is required to realize adequate trust and involvement of users. This paper presents the construction of a Bayesian network model for CTL as refinement of a situated theory on naval operators' information processes.

1.1 Cognitive Task Load

The cognitive task load (CTL) theory distinguishes three load dimensions. The first dimension is the *time occupied*, which is high when the operator has to work with maximum cognitive processing speed to search and compare known visual symbols or patterns, to perform simple (decision-making) tasks, and to manipulate and deal with numbers in a fast and accurate way. With respect to the second dimension, the *level of information processing*, (a) information that is processed automatically, results into actions that are hardly cognitively demanding, (b) routine procedures involve rather efficient information processing, and (c) problem solving and action planning for relatively new situations involve a heavy load on the limited capacity of working memory. *Task-Set Switches* is the third load dimenssion, addressing the demands of attention shifts or divergences in which different sources of human task knowledge have to be activated. It should be noted that the effects of cognitive task load depend on the concerning task duration. In general, the negative effects of under- and overload increase over time.

1.2 Emotional State

Neerincx [4] proposes to combine the CTL-model with a model of the Emotional State (ES) for high-demand task domains in which the human sometimes works in extreme and critical conditions. The ES-model distinguishes two dimensions: the arousal level—low versus high—and the valence level—positive versus negative [5]. Emotion and CTL are related: for specific load conditions a specific emotional state ("response") can be expected. For example, when task load increases, an adequate response is to invest extra effort (i.e., arousal increases) in order to maintain good performance [3].

1.3 Model Levels

For the CTL-ES model, we distinguish three levels (Fig 1). The first level describes the *human act observables*, which are behavioral and bodily variables that correlate with human information processes (HIP).

At the second level, *HIP dimensions* represent variables that correlate with human performance. SOWAT, an activity monitoring tool, can be used to derive the CTL-dimensions' values from observables as user-interface acts [2], while affective computing techniques can be used to derive the ES-dimensions' values from, for example, facial and speech expressions [6]. An operator profile can be applied for personalized estimation of HIP-dimensions' values from observables. For example, the level of experience influences the Level of Information Processing (LIP): the higher the experience, the lower the LIP value. The dimensional model is trained in advance by datasets that include performance measures. This estimation may concern the current performance and the near-future performance.

At the third level, *HIP classes* are derived from the dimensional models. CTL-classes are underload (UL), overload (OL), vigilance (VI), cognitive lock-up (CL), and neutral (NE); ES-classes are boredom (BO), relaxed (RE), excited (EX), stressed (ST), and neutral (NE).

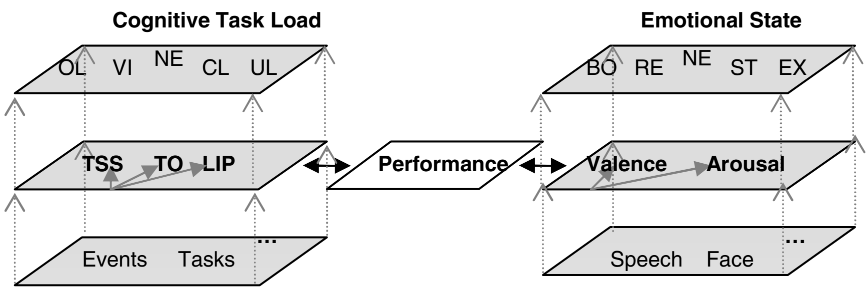


Fig. 1. The Performance, Cognitive Task Load and Emotional State model

1.4 Performance Estimation

This paper focuses on the construction of the dimensional CTL-model (i.e., the 2nd level of Fig. 1). For this purpose, we need a method to analyze data from training and actual task performances, which can cope with missing data. Furthermore, it should be easy to extend the model, for example, starting with CTL-dimensions and adding ES-dimensions when appropriate. In addition, the model should be transparent (i.e., providing a structure that gives insight in which variable influences other variables), enabling estimations of near-future values. Bayesian networks seem to fulfill these requirements. This paper investigates whether a Bayesian network can be constructed that provides adequate estimations of the CTL-performance relationships for two datasets: operator performance on a high-tech sailing ship and operator performance with an adaptive user interface in a lab-setting.

2 Bayesian Networks

Bayesian networks are graphical models for reasoning under uncertainty. A Bayesian network consists of a network structure and conditional probability tables. The structure of a Bayesian network consists of nodes and arcs. The nodes represent variables, and the arcs represent direct dependencies between the variables. If there is an arc from one node to another, then the first node is called the parent of the latter (the child). The structure of a Bayesian network is a directed acyclic graph (DAG). In other words, the structure does not contain any cycles. Each node has a conditional probability table. This table defines the probabilities of that node on taking each of its values, given its parent(s). Bayesian networks are often applied in the medical domain. Given symptoms, the Bayesian network can compute the probability of the presence of a disease using Bayes' Theorem (see Equation 1).

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \tag{1}$$

This is called Bayesian inference and can be explained with the hypothetical network structure of Fig. 2d as example, in which performance relates to TO, TSS and LIP. Table 1 shows possible conditional probability tables. If there is evidence that a certain person has low performance, the probability that this person experiences high TO, TSS and LIP can be read in the tables. These probabilities are respectively 0.5, 0.4 and 0.6. The other way around, it is possible to calculate the probability that a person has a low performance when high TO, TSS and LIP are observed. This can be done using Bayes' Theorem:

$$\begin{aligned}
&P(\text{low performance with high TO, TSS and LIP})\\
&= P(\text{PERF} = l \mid \text{TO} = h \wedge \text{TSS} = h \wedge \text{LIP} = h)\\
&= \frac{P(\text{TO} = h \wedge \text{TSS} = h \wedge \text{LIP} = h \mid \text{PERF} = l) \cdot P(\text{PERF} = l)}{P(\text{TO} = h \wedge \text{TSS} = h \wedge \text{LIP} = h)}\\
&= \frac{P(\text{TO} = h \wedge \text{TSS} = h \wedge \text{LIP} = h \mid \text{PERF} = l) \cdot P(\text{PERF} = l)}{\sum_{p \in \{l,m,h\}} P(\text{TO} = h \wedge \text{TSS} = h \wedge \text{LIP} = h) \mid P(\text{PERF} = p)}\\
&= \frac{0.5 \cdot 0.4 \cdot 0.6 \cdot 0.3}{0.5 \cdot 0.4 \cdot 0.6 \cdot 0.3 + 0.5 \cdot 0.1 \cdot 0.5 \cdot 0.4 + 0.1 \cdot 0.0 \cdot 0.1 \cdot 0.3}\\
&= \frac{0.036}{0.036 + 0.01 + 0}\\
&= 0.783
\end{aligned}$$

Table 1. Possible conditional probability tables for the network structure of Fig. 2d

Performance		
low	medium	high
0.3	0.4	0.3

	TSS		
Performance	low	medium	high
low	0.3	0.3	0.4
medium	0.6	0.3	0.1
high	0.6	0.4	0.0

	TO		
Performance	low	medium	high
low	0.0	0.5	0.5
medium	0.1	0.4	0.5
high	0.7	0.2	0.1

	LIP		
Performance	low	medium	high
low	0.0	0.4	0.6
medium	0.1	0.4	0.5
high	0.7	0.2	0.1

3 Experiment: Analysis of Two Datasets

To create a Bayesian Network for Performance and Cognitive Task Load, we analyzed two datasets: the first dataset was automatically collected during operator's interaction with a prototype user interface, and the second dataset was manually collected during operator's performance on a sailing ship.

3.1 Lab Dataset

The Lab data were acquired during an experiment at the MBO Shipping & Transportation College of Rotterdam (for details, see [7]). 12 students participated, all second and third year students (average age of 20.1 with a standard deviation of 2.1, 11 males, 1 female; relevant knowledge about the maritime domain). All participants had to deal with alarms during platform supervision, damage control and navigation tasks. All performed actions were recorded in log files and used to calculate TO, TSS, LIP and performance, with use of SOWAT [2].

The Lab data contained 1407 cases with data for LIP, TSS, TO and Performance. Each case in the data file corresponds to a sliding window of 60 seconds with 50 seconds overlap. The values for LIP range from 0 (low) to 6.5 (high), TSS ranges from 0 to 5, TO ranges from 0% to 100%, and performance ranges form 0 (low) to 4 (high). All values of the variables were converted to the values low, medium and high for our analyses. Since Bayesian networks are best trained with data that have an equal distribution, we have chosen the thresholds to accomplish this as much as possible (see Table 2 for the distribution).

From this data file we created a balanced train and test set. We have selected 333 cases with low performance randomly form the total of 427 cases with low performance, and did the same for medium and high performance. The test set contained 150 cases, also with an equal distribution that was randomly selected. The other 258 cases were not used for training or testing since this would result in unbalanced train and test sets.

Table 2. Distribution of cases over CTL and Performance for the two datasets

	Lab data				Ship data			
	TO	TSS	LIP	Perf.	TO	TSS	LIP	Perf.
Low	476	722	462	427	571	1123	426	373
Medium	460	425	468	398	599	378	591	390
High	471	260	477	582	582	251	735	989

3.2 Ship Dataset

The Ship data were acquired during an experiment in the Ship Control Centers of three sailing air defense and command frigates (for details, see [8]). Each ship was manned with four active duty teams, data collection concerned two persons of each team. In total there were 12 teams and 24 participants (all male). Each team had to perform three scenarios that varied in TO, TSS and LIP. All scenarios were recorded on video and scored by experts afterwards on TO and LIP. LIP was scored by the participants themselves. SOWAT [2] was used for integration of all data and generation of 1752 cases. Each case in the data file corresponds to a sliding window of 60 seconds with 40 seconds overlap. The values for LIP range form 1 (low) to 5 (high), TSS ranges from 0 to 6, TO ranges from 0% to 100%, and performance ranges form 0 (incorrect or too slow response) to 2 (correct response). All values of the variables were converted to the values low, medium and high. For this dataset we have also chosen the thresholds to accomplish an equal distribution as much as possible (see Table 2 for the distribution).

From this data file we created a train set and a test set. The train set contained 969 cases with an equal distribution of performance. The cases were also randomly selected. The test set contained 150 cases, also with an equal distribution that was randomly selected. The other 633 cases were not used for training or testing.

3.3 Creating the Network Structure

When creating a Bayesian network, the structure of the network can either be defined by an expert, or learned from a dataset. We used GeNIe 2.0[1] to create four network structures for each dataset. GeNIe is equipped with four structure learning algorithms:

- Essential Graph Search (EGS) algorithm [9]
- PC algorithm [10]
- Greedy Thick Thinning (GTT) algorithm
- Naïve Bayesian network (NBN) algorithm [11]

After creating the network structures we created the conditional probability tables using Netica-J's[2] parameter learning algorithm. This algorithm was applied to the same train sets that were used for structure learning.

Finally, the performance of the created Bayesian networks was tested with the test sets using Netica-J's performance testing algorithm. These results were evaluated using a Chi-square test.

3.4 Results

This section first shows the results for the Lab and Ship datasets, then discusses the generalizability of the networks.

3.4.1 Lab Data

The network structures that were created by the four structure learning algorithms using the Lab train set are, with the exception of the NBN algorithm, very similar. The first three algorithm produce a fully connected network structure, the only difference is the direction of the arcs (Fig. 2).

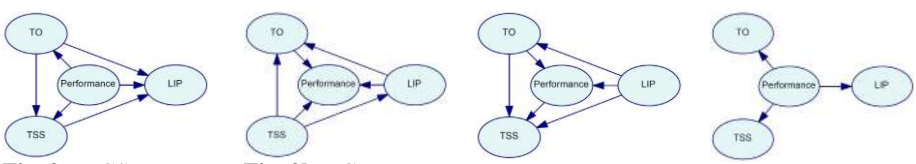


Fig. 2a. EGS **Fig. 2b.** PC **Fig. 2c.** GTT **Fig. 2d.** NBN

Fig. 2. The created network structures for the Lab dataset using the four algorithms

[1] http://genie.sis.pitt.edu/
[2] http://www.norsys.com/

As a result of the similarity in the network structure of Fig. 2, the percentages of cases classified correctly are the same for these three algorithms. All four algorithms perform overall significantly better than random guessing a performance level (all p<0.000). The NBN algorithm performed overall slightly better than the other algorithms, but this difference was not significant (p<0.816). When we zoom in to the different performance categories, we see that de difference between random guessing and the NBN algorithm is significant for low performance (p<0.014). For medium performance, the difference between the EGS, PC and GTT algorithms and random guessing are significant (p<0.008). Finally, for high performance the difference between the four algorithms and random guessing is significant (p<0.001).

When we look at the network with the highest performance in detail, we see that it is not able to distinguishing well between low and medium performance (see Table 3, left). When we join the performance categories low and medium together, the percentage correct classified increased from 58% to 85% (see table 3, middle), while the expectation value ("random") increased to from 33% to 50%. A drawback of this method is that the dataset is not distributed equally for performance. To accomplish an equal distribution we adjusted the threshold for performance. The Bayesian network was trained with a train set that consisted of 500 cases with low and 500 cases with high performance. The network was tested with a test set that contained 50 cases with low and 50 cases with high performance. This network classified 86% of the cases correct. More importantly, all cases with low performance were recognized, see Table 3, right. This Table is the same for all network structures that were tested. In other words, all network structures perform the same, see Fig. 3, right.

Table 3. Performance of the networks with the highest percentage correct classified with three (left) and two performance levels, unbalanced (middle) and balanced (right)

	Prediction		
Actual	low	medium	high
low	29	21	0
medium	20	23	7
high	10	5	35

	Prediction	
Actual	low	high
low	93	7
high	15	35

	Prediction	
Actual	low	high
low	50	0
high	14	36

The networks that were trained with two performance categories performed overall better than the networks that were trained with three performance categories, even after correction for chance using Cohen's Kappa.

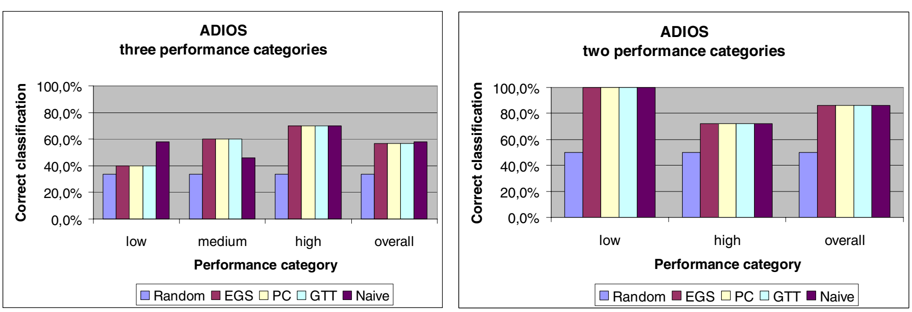


Fig. 3. Network performance of the different algorithms for the for the Lab dataset with three (left) and two (right) performance levels (balanced)

3.4.2 Ship Data

The network structures that were created by the four structure learning algorithms using the Ship train set show more variation (Fig. 4) than we have seen with the Lab dataset (Fig. 2). The structures are not fully connected and with the exception of the NBN algorithm, there is no direct dependence between TSS and performance.

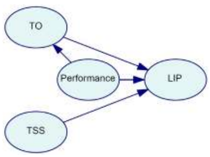


Fig. 4a. EGS

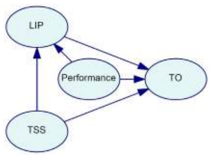


Fig. 4b. PC

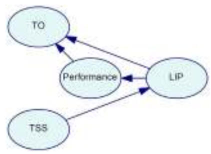


Fig. 4c. GTT

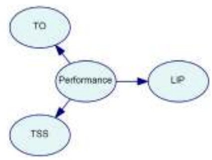


Fig. 4d. NBN

Fig. 4. The created network structures for the Ship dataset using the four algorithms

As a result of the variation in network structure, the percentages of cases classified correctly differ considerable (Fig. 5, left). All four algorithms perform overall significantly better than random guessing a performance level ($p<0.000$ for the EGS and PC algorithm, $p<0.005$ for the GTT algorithm and $p<0.045$ for the NBN algorithm). The PC algorithm shows the best performance, but does only perform significantly better than the NBN algorithm ($p<0.029$).

The Ship dataset was also tested with two performance levels (Fig. 5, right). The percentage classified correct of the best network increased from 57% to 76%, while the expectation value ("random") increased to from 33% to 50%.

The networks that were trained with two performance categories performed overall better than the networks that were trained with three performance categories, even after correction for chance using Cohen's Kappa.

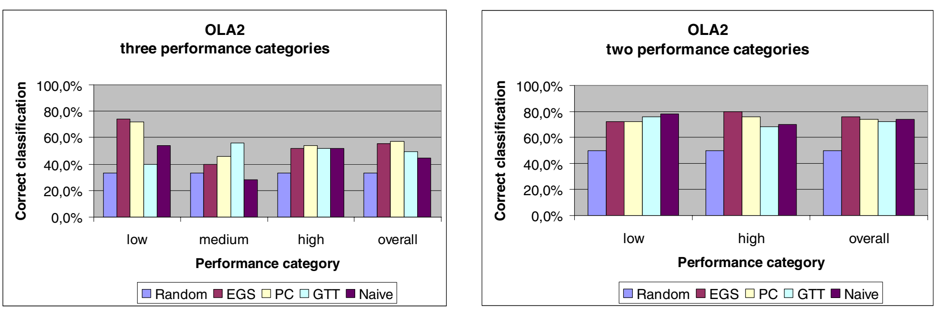


Fig. 5. Network performance of the different algorithms for the for the Ship datasets with three (left) and two (right) performance categories (balanced)

3.5 Generalizability

To test the generalizability of the networks, we tested the performance of the networks that were trained with the Ship train-set with the Lab test-set and vice versa (Table 4). When these results are compared with the results of the networks that have

Table 4. Cross dataset testing

		Correct classification (%)							
		3 performance categories				2 performance categories			
Train set	Test set	EGS	PC	GTT	NBN	EGS	PC	GTT	NBN
Lab	Lab	56.7	56.7	56.7	58.0	86.0	86.0	86.0	86.0
Ship	Lab	56.0	53.3	54.0	56.0	79.0	81.0	84.0	84.0
Lab	Ship	56.7	56.7	56.7	58.0	70.0	63.0	71.0	74.0
Ship	Ship	55.3	57.3	49.3	44.7	76.0	74.0	72.0	74.0

been tested with the same datasets as they were trained, we see that almost all differences are not significant. The only exception is the network that was created with the PC algorithm using the Lab data with two performance categories, and tested with the Ship data ($p<0.009$).

4 Conclusions and Discussion

Previous research showed the effects of CTL on operator task performance, and possible mitigation methods (adaptive task allocation, cognitive aids and resource feedback). This paper provides the first results on applying Bayesian Networks to model these effects in order to estimate and predict possible performance shortcomings. We derived the CTL-performance relationships for two datasets: operator performance with an adaptive user interface in a lab-setting and operator performance on a high-tech sailing ship (Ship). The first dataset provides the best results, probably because the recording was conducted in rather controlled conditions and all three CTL-factors showed variance in the scenario. In contrast, the dataset of the sailing ships contained relatively few Task-Set Switches (TSS), which might explain the creation of network structures that do not include a direct relationship of TSS with Performance (see Fig. 4). However, the “Naïve Bayesian Network” model that is trained with the more-balanced Lab dataset proves to provide similar performance prediction results for the Ship dataset as the models that are derived from the Ship training dataset (i.e., for the two category performance, see Table 4). So, the “Naïve Bayesian Network” algorithm seems to be a good choice, providing performance estimations with 86% and 74% accuracy for respectively the lab-setting and sailing ship data (with respectively a 100% and 78% hit-rate for the low performance category). Overall, the resulting model nicely generalizes over the two datasets. Although the results are relatively positive, there is a clear room for improvement. Currently, we are extending the modeling approach with emotion, both for the defense and the space domain. A major question is how to adequately address the occurrence of very rare cases for which the dataset is not trained? A method to detect such occurrences would be very beneficial.

References

1. Neerincx, M.A.: Cognitive task load design: model, methods and examples. In: Hollnagel, E. (ed.) Handbook of Cognitive Task Design, ch. 13, pp. 283–305. Lawrence Erlbaum Associates, Mahwah (2003)

2. Grootjen, M., Neerincx, M.A., Stolk, K.D., van Weert, J.C.M., Bierman, E.P.B.: Design and user evaluation of an interface prototype that adapts to the operator's cognitive task load. In: Schmorrow, D.D., Nicholson, D.M., Drexler, J.M., Reeves, L.M. (eds.) Proc. 4th Intern. Augmented Cognition, pp. 97–106. Strategic Analysis, Inc., Arlington (2007)
3. Neerincx, M.A., Bos, A., Olmedo-Soler, A., Brauer, U., Breebaart, L., Smets, N., Lindenberg, J., Grant, T., Wolff, M.: The Mission Execution Crew Assistant: Improving Human-Machine Team Resilience for Long Duration Missions. In: Proc. of the 59th International Astronautical Congress (IAC2008), 12 pages. IAF, Paris (2008), DVD: ISSN 1995-6258
4. Neerincx, M.A.: Modelling Cognitive and Affective Load for the Design of Human-Machine Collaboration. In: Harris, D. (ed.) HCII 2007 and EPCE 2007. LNCS, vol. 4562, pp. 568–574. Springer, Heidelberg (2007)
5. Truong, K.P., van Leeuwen, D.A., Neerincx, M.A.: Unobtrusive Multimodal Emotion Detection in Adaptive Interfaces: Speech and Facial Expressions. In: Schmorrow, D.D., Reeves, L.M. (eds.) HCII 2007 and FAC 2007. LNCS (LNAI), vol. 4565, pp. 354–363. Springer, Heidelberg (2007)
6. Bradley, M., Lang, P.: Measuring emotion: The Self-Assessment Manikin and the Semantic Differential. Journal of Behavioral Therapy & Experimental Psychiatry 25, 49–59 (1994)
7. Grootjen, M., Reijenga, Y., Neerincx, M.A.: A user evaluation of an interface that adapts to the user's cognitive task load (to appear)
8. Grootjen, M., de Greef, T., Neerincx, M.A.: Effects of Level of Automation on Operator Task Load and Performance on a Sailing Naval Ship (to appear)
9. Dash, D., Druzdzel, M.: A hybrid anytime algorithm for the construction of causal models from sparse data. In: Proc. 15th Annual Conference on Uncertainty in Artificial Intelligence, pp. 142–149. Morgan Kaufmann, San Francisco (1999)
10. Spirtes, P., Glymour, C.N., Scheines, R., Causation: Prediction, and Search. MIT Press, Cambridge (2000)
11. Duda, R.O., Hart, P.E.: Pattern Classification and Scene Analysis. Wiley, New York (1973)

Impact on Performance and Process by a Social Annotation System: A Social Reading Experiment

Les Nelson, Gregorio Convertino, Peter Pirolli, Lichan Hong, and Ed H. Chi

Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304
{lnelson,convertino,pirolli,hong,echi}@parc.com

Abstract. Social annotation systems such as SparTag.us and del.icio.us have been designed to encourage individual reading and marking behaviors that, when shared, accumulate to build collective knowledge spaces. Prior work reported on the experimental design and performance effects observed in a controlled study of SparTag.us. Study participants working independently on a sensemaking task who had access to a set of expert annotations were compared against participants using SparTag.us without those annotations and participants using only office software for annotation support. A learning effect favored the participants exposed to expert annotations. In this paper, we analyze the behavioral data captured during the experiment and identify differences in the work process that can explain the performance effects reported previously.

Keywords: Convergent measures, social annotation systems, evaluation, social sensemaking.

1 Introduction and Approach

Learning and knowledge handoff are becoming critical in the workplace. In fact, knowledge work is increasingly depended on (or equivalent to) learning and professional development (Tapscott 1996). Also, corporations are often losing critical domain knowledge as older workers are leaving before they could transfer their knowledge.

But new technologies are showing new opportunities. Web 2.0 tools have lowered the costs for social construction of knowledge (e.g., Wikipedia, social bookmarking) and made possible user-defined combinations of content across web services (e.g., mashups). New semantic web techniques are also allowing users to give structure to the content that they share (e.g., microformats in blogs).

In this context of new social needs and social technology, researchers in Human-Computer Interaction and Information Retrieval are redirecting their focus of inquiry from solitary individuals working with systems and content to models of social information foraging, knowledge sharing, and sensemaking [7, 11]. Social annotation

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 270–278, 2009.

systems such as SparTag.us [2] and del.icio.us[1] exemplify socially constructed understandings of content.

However, the settings in which collaborative software is employed are full of experimental confounds: real-world socio-technical systems introduce greater complexity into the evaluation process [2]. Measures of performance in social sensemaking remain difficult [4]. Researchers of communication and collaboration technologies have faced this problem repeatedly. As a result, they have started including also process measures (e.g., measures of costs in the process such as turn-based measures of efficiency in communication). Monk and collaborators [5] reconstruct the evolution of measures in these studies. They observe that the traditional measures of task performance such as number of errors and completion time were only sensitive to gross changes in the technology utilized. For example, when performing tasks within experiments, the participants may tend to protect their primary task and get the work done efficiently through extra effort (i.e., costs) at the expense of any secondary tasks. In order to capture these hidden effects the researchers have introduced measures that characterize aspects of the process of communication, rather than just the final outcomes (e.g., number of the turns, length of the turns, kind of turns, see [9]). More recently, experiments on knowledge sharing in teams have combined both process and performance measures in order to assess the effects of the new tool on the sharing process (quality and costs) and then the consequences of these effects on the group performance [e.g., 1].

Previously we reported the main results obtained from performance measures on the use of the social annotation tool, SparTag.us [6]. This show a statistically significant increase in subject matter learning for participants using the tool in a condition of access to annotations of another. In this paper we briefly summarize the method and the effects on performance and then focus on the process measures that were also taken during the experiment. The goal is to use behavioral measures (e.g., URLs visited) and supplemental products (e.g., responses to essay questions) to help characterize visible changes in the process that led to the differences in performance (questionnaire-based outcome). The analysis of these process measures is in part quantitative and in part qualitative.

We next briefly introduce the object of study, the social annotation tool, SparTag.us. We then summarize the experiment and its measures and detail selected process measures found. We discuss these in light of the significant performance gain seen in subject matter learning. We conclude with implications for design and future research.

2 The Study System, SparTag.us

Inspired by the work of Schraefel et al. [8] showing that in many cases of information foraging the content of interest is at the sub-document level, SparTag.us uses annotation as a means to collect paragraphs of interest. Specifically, when a user loads a web page in his browser, we modify the underlying representation of the page to partition its textual content into paragraphs and make the words of the paragraphs live and

[1] http://del.icio.us

clickable [3]. Here the user can annotate the content in various ways. S/he can click on words of a paragraph to tag the paragraph, This Click2Tag interface offers a low-cost option for the user to annotate paragraphs of interest in situ while reading the web page. S/he can also highlight phrases and sentences in situ through click-and-drag actions. Thirdly, as the user tags or highlights, SparTag.us automatically extracts the annotated paragraphs from the page and inserts them into a system-created notebook, where further annotations can be made later.

In SparTag.us, a user can also subscribe to the annotations of another user by designating that user as a friend. Consequently, the user will see his friend's annotations when viewing pages containing the same paragraphs annotated by his friend. Color-coding is used to distinguish between own and friends' annotations. Figure 1 shows the friend's notebook as viewed by the user. Note that the friend's highlights and tags are displayed in light blue and the user's own highlights in yellow.

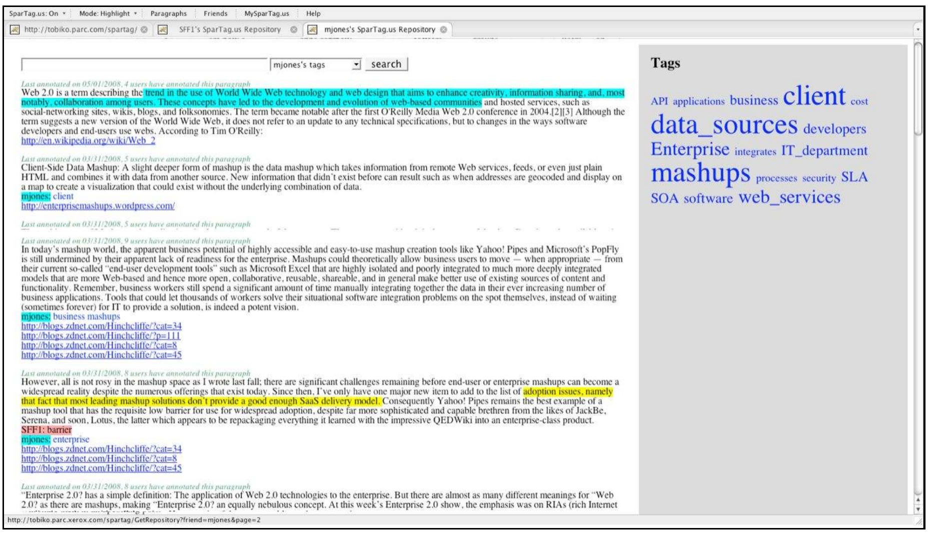


Fig. 1. Study participants using the SparTag.us annotation tool may view the collected tags and highlights of another in a friend's Notebook. The view here shows annotated paragraph with another's annotations shown in blue and the other person's tag cloud.

3 A Social Reading Experiment

We conducted a 'Social Reading Experiment' where participants needed to use Web resources to learn about a topic area: "Enterprise 2.0 Mashups", which is a combination of the technology areas of "Enterprise 2.0"[2] and "Web 2.0 Mashups". Study participants would need to find and understand many web pages because at the time of the study there was no single source of information on the topic area.

[2] http://en.wikipedia.org/wiki/Enterprise_2.0

Our experiment compared three groups of participants who worked:

1. Without SparTag.us (WS), but with traditional note-taking tools.
2. With SparTag.us only, used individually (SO).
3. With SparTag.us with the annotations of a ‘Friend’ (SF).

The conditions WS and SO were control conditions in which individuals read web content without access to others’ annotations. To provide for an ecologically valid comparison, WS participants could take notes in MS Word or with pen and paper. In the SF condition, people independently read web content but also had access to social annotations created by an experimenter-simulated subject-matter expert.

Tools like SparTag.us and del.icio.us are tools used at an Internet scale and scope. In our experimental setup we look at the performance of individual users. However, we extended the scope of inquiry beyond the individual by simulating a social reading condition. That is, in one of the conditions each user was exposed to the SparTag.us Friend, which is an organized collection of annotations comprising a tag cloud, a list of URLs, and a set of paragraphs. These annotations are derived from the following social resources. Twenty tags associated with the top 100 annotated URLs from a del.icio.us query on “enterprise mashup” constituted the target tag cloud. URLs found by top hits from a Google search that used each tag as a search term. These URLS were manually tagged with these ‘expert’ tags using SparTag.us.

The hypothesis is that participants that were exposed to tags, URLs, and highlights from a knowledgeable other would perform better than the participants without this exposure. We thus evaluate performance measures between subjects in the experimental condition, SF, with those in control conditions, SO and WS. Eighteen participants completed two experimental sessions. The first day was a four hour series of demographic survey, true-false question answering, learning in the domain area lasting two hours, one writing essay, and a debrief. Day 2 lasted one hour and involved one true-false question set and a second writing task. More details on the procedure can be found in [6].

We used a combination of performance and process measures to understand the impact of the annotation support used, but also give indications of how people are employing the technology in the context of their reading and annotation practices. The performance was measured using a questionnaire (created for this study). The questionnaire included a set of true-false questions, which were generated from an expert elicitation process and were used to assess objective learning gains in the subject matter domain before and after the users foraged the information in each of the three conditions.

The process measures pertained to the reading and writing behaviors of each participant: the number and sequence of Web resources visited (logged by Universal Resource Locator or URL), loaded and scrolled; the annotations made (tags and keywords used), and the personal notes taken during the task.

The main measure of learning (equation (1)) was obtained through a metric of learning effect developed as part of the experimental method. The Gain metric is a composite indicator that was computed on the basis of several scores derived from the questionnaire: Pretest to Posttest questionnaire scores for each participant, and maximum score. Specifically, gain scores were calculated as:

$$Gain = \frac{(PostTest\_Score - PreTest\_Score)}{(Max\_Score - PreTest\_Score)} \tag{1}$$

Using the Gain metric as the measure of learning performance, we report in [6] a learning effect, with the SF group showing significantly greater gains than the SO group and the WS group. The WS and SO groups were not significantly different.

This establishes that participants with access to resources from a knowledgeable other exhibited a greater learning performance. What might have caused this effect? What costs are reduced or what kinds of benefits are increased? In the remainder of this paper we turn to the other measures taken during this experiment to explore these questions.

4 Results of Process Measures of Reading Activities

We previously reported in [6] differences in reading activities amongst the groups. While not statistically significant there was a consistent trend seen that on average SF participants visited fewer URLs, but spent more reading time on those they visited. This suggests a more in-depth analysis of the fewer sources of information chosen by SF participants.

Table 1. Trends suggest different reading behaviors between conditions

	URL Visits		Time on URL	
Group	Mean	SD	Mean (sec)	SD
SF	59	23.7	144.5	73.0
SO	71.2	25.5	128.2	56.1
WS	79.3	35.9	87.3	24.0

We look further into this by examining what kind of URLs were being visited. Sites were classified as representing the following kind of information sources:

- Blog, indicating the site was an individual's Web log;
- Conference, an industry or academic conference site;
- Consultant, the business site of a consulting service;
- Employment, a job posting site;
- MySpartagus, use of the SparTag.us Noteboook;
- News, a general or technology news service;
- OpenSource, information site of the Open Source community;
- Search, an Internet Search service;
- Vendor, a site of a business selling in the domain area;
- Wikipedia.

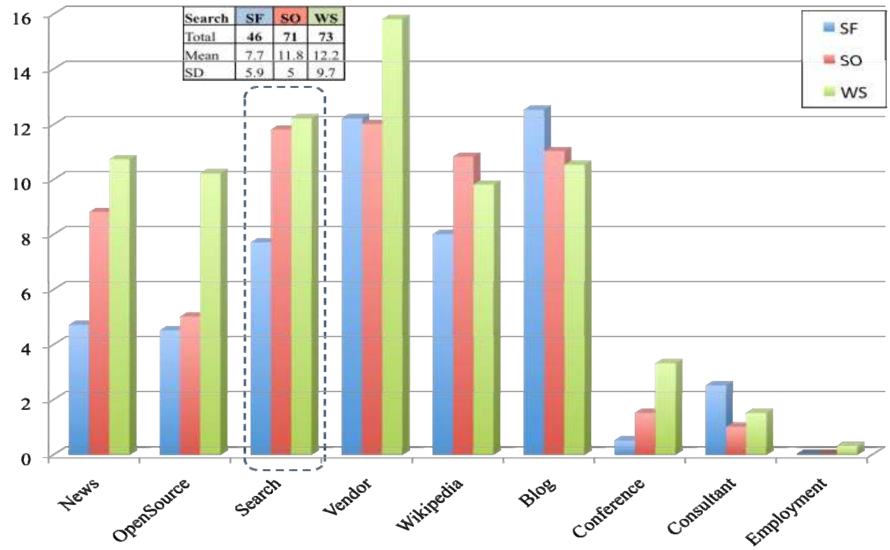


Fig. 2. Counts of URLs visits by kind of source while Reading. Search actions are fewer for SF than for SO and WS participants.

Our aim here is to look at the differences in Web resources used to discover traces of cost/benefit differences between the study groups. Figure 1 shows the classification of the 1149 web sites visited by the 18 participants during the reading portions of the experiment. We can see that search is a main resource (16.5% of all URL visited) and that search behavior exhibits the same trend of distinguishing SF against SO and WS in terms of lower use of search during the observed information foraging.

5 Results of Process Measures of Writing Activities

We previously reported [6] differences in writing activities amongst the groups. While not statistically significant there was a consistent trend seen that on average SF participants used more domain terminology in writing answers to their essay questions (see Table 2).

Table 2. People using SparTag.us used more domain words

Group	All Words	SD	Domain Words	SD
SF	549.92	207.01	141.50	46.27
SO	528.92	202.08	136.00	58.68
WS	459.67	174.32	117.00	48.17

Similar to the previous section we used a classification of the kind of URLs visited during writing activities. We looked at only new Web resources that were used during the writing tasks to supplement those found during the prior reading activity. This includes far fewer URLs visited, where ‘Consultant’ Web sites were not visited and new kinds were seen (i.e., private library catalogs and pages from online copies of published books). Table 3 shows the collected use of new Web resources during writing. We again see a trend indicating need for less information finding amongst SF participants over the other groups.

Table 3. Participants in the SF condition accessed fewer supplemental Web resources to answer the essay questions

Group	Total	M	SD
SF	39	6.5	5.1
SO	58	9.7	7.7
WS	60	10.0	12.6

6 The Devil Is in the (Process) Details

We had found that users supported by our social annotation tool and having access to annotations from a domain expert (i.e., SF condition) showed a significant increment in subject matter learning. In this study we addressed the question of ‘how’ (i.e., ‘in what ways’) such improvement in performance had occurred. To this end, we analyzed process measures characterizing the foraging behavior both before and during the writing of the report (i.e., number of URLs read, average time spent per URL, kind of URL, number of searches). We found that compared to the other two conditions the participants in the SF condition exhibited greater efficiency in foraging (i.e., fewer sources visited, more time per source, fewer searches) and greater efficiency in producing the final report (i.e., more words written in the same with less additional foraging activity done while writing).

Given the experimental differences imposed among the conditions (WS, SO, SF) and the abovementioned better learning performance of the SF participants, these results about the process suggest that having access to the annotations from a domain expert reduced the costs of foraging information, promoted more focus and depth of analysis, and saved time that the SF participants used to write content in the report.

These results point to directions for future work. A more detailed explanation of the how and why these effects occurred will help us understand how they could be induced in other situations (e.g., in peer-to-peer group collaboration or within communities of practice). More detailed exploration of the stimuli used in the interventions is needed.

The SO condition (with SparTag.us only) affords collecting relevant paragraphs in a notebook and, for each paragraph, highlighting relevant sentences, or labeling it with any of its own word (click-to-tag) or with a new user-entered keyword. The SF condition was, by design, the condition with highest support for learning because it further included structured information including three kinds of stimuli:

1. A cloud of tags that represented the expert's terms for the domain.
2. A set of URL to jumpstart the foraging process from relevant sources.
3. A set of sample paragraphs in the expert's (or friend's) notebook which were examples of pieces of relevant information that the user could expect to find and then annotate in her/his notebook.

As part of our future work we plan to examine in detail which of these three sets of stimuli has more effects, at what stage of the process, and potential interactions between them. This requires focused follow-up studies that adopt measures consistent with the present study. These could manipulate solely the exposure to the expert's tags and then measure the effects on the terms typed while foraging new information; or could manipulate only the exposure to the expert's set of URLs and measure the volume and ordering of the sources read and annotated in the user's notebook; or, finally, could manipulate the visibility of sample paragraphs in the friend's Notebook and measure if the final reports by users exposed are individually more focused on fewer topics and/or more consistent among each other (e.g., measure, within each condition, how similar the content of the report is to the content of the paragraphs and/or the cloud tags). In summary, what beneficial effects do experts' tags, URLs, and relevant paragraphs have? How and at what stage of the process are learners influenced by these expert traces? What are the extra cognitive costs when these cues are missing?

Acknowledgments

Research funded by the ONR grant No. N00014-08-C-0029 to Peter Pirolli. We thank Chirstoph Held and Diane Schiano for their assistance in designing and conducting this study.

References

1. Convertino, G., Mentis, H.M., Rosson, M.B., Carroll, J.M., Slavkovic, A., Ganoe, C.H.: Articulating common ground in cooåperative work: content and process. In: Proc. CHI 2008, pp. 1637–1646. ACM, New York (2008)
2. Grudin, J.: Groupware and social dynamics: Eight challenges for developers. CACM 37(1), 92–105 (1994)
3. Hong, L., Chi, E.H., Budiu, R., Pirolli, P., Nelson, L.: SparTag.us: A low cost tagging system for foraging of web content. In: Proc. AVI 2008, pp. 65–72. ACM, New York (2008)
4. Kalnikaite, V., Whittaker, S.: Social summarization: Does social feedback improve access to speech data? In: Proc. CSCW 2008. ACM, New York (to appear, 2008)
5. Monk, A., McCarthy, J., Watts, L., Daly-Jones, O.: Measures of Process. In: CSCW requirements and evaluation, ch. 9, pp. 125–139. Springer, Berlin (1996)
6. Nelson, L., Held, C., Pirolli, P.L., Hong, L., Schiano, D.J., Chi, E.H.: With a Little Help from My Friends: Examining the Impact of Social Annotations in Sensemaking Tasks. In: Proc. CHI 2009 (2009)
7. Pirolli, P.: A Multilevel Science of Social Information Foraging and Sensemaking. In: Position Paper, Information Seeking Support Systems Workshop, Sponsored by the US National Science Foundation, Chapel Hill, NC USA (June 2008)

8. Schraefel, M., Zhu, Y., Modjeska, D., Wigdor, D., Zhao, S.: Hunter Gatherer: Interaction Support for the Creation and Management of Within-Web-Page Collections. In: Proc. WWW 2002, pp. 172–181 (2002)
9. Sellen, A.J.: Remote conversations: the effects of mediating talk with technology. Human-Computer Interaction 10, 401–444 (1995)
10. Tapscott, D.: The Digital Economy: Promise and Peril in the Age of Networked Intelligence. McGraw-Hill, New York (1996)
11. Vicente, K.J.: HCI in the global knowledge-based economy: designing to support worker adaptation. ACM Transactions on Computer-Human Interaction 7(2), 263–280 (2000)

Proposing Strategies to Prevent the Human Error in Automated Industrial Environments

José A. do N. Neto[1], Maria F.Q. Vieira[1], Charles Santoni[2], and Daniel Scherer[1]

[1] LIHM –DEE – CEEI UFCG, Campina Grande, Brazil
[2] LSIS, Université Paul Cézanne (Aix-Marseille III), France
{fatima,alves,daniel}@dee.ufcg.edu.br,
charles.santoni@lsis.org

Abstract. This paper presents a process to conceive strategies to prevent the human error when operating industrial systems. The process adopts a broader view to error prevention, going beyond the error analysis to consider the user profile, the task and context description. The error classification is done according to a task execution cognitive model. The conceived strategies focus on the human interface component of those systems since it is this work's premise that the human interface design has a strong impact on the human error rate.

Keywords: Human-Machine Interfaces, Human error, Cognitive model.

1 Introduction

With the technology development and the consequent rise in system robustness it becomes more common the occurrence of incidents and accidents related to human errors. In parallel, automated systems concentrate information and decision making on the hands of fewer operators who under time pressure are subjected to high cognitive loads. In the electricity industry, as it will be discussed later in the case study, 20% of system failures are related to human errors. This work is part of a broader study that aims to build more ergonomic human interfaces for automated industrial systems. Those systems operate efficiently during routine but are highly dependant on the human operator during exception and critical situations, when tasks are more complex and deadlines more strict.

This paper presents a process to conceive strategies to prevent the human error when operating industrial systems. The process is based on the error classification according to a task execution cognitive model. The strategies are conceived for the human interface component of the automated systems since it is this work's premise that the human interface design has a strong impact on the human error rate.

The process is based on a broad context analysis that includes: work context, operator profile, task analysis and the history of incidents and accidents that are related to the human error in the industrial installation. The operator profile brings information on the human abilities and limitations, thus being essential to understand the causes of errors and how to prevent these. The task analysis, beyond describing task characteristics, must also give insight into task frequency, the impact of the error and

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 279–288, 2009.

the interaction rules for the system operating devices. The context analysis is based on ergonomic principles and consists on detailing the work environment from the physical and organizational points of view. It must highlight issues related to work ergonomics and their potential relation with system failures due to human error. To complete this phase, there is the analysis of the history of system failures related to human error, according to reports on accidents and incidents. The errors are then classified according to the categories proposed in the work of Rouse and Rouse, apud Cellier [2].

The purpose of this broad analysis is to identify the relationship between the human errors and the quality of the human interface; in order to propose effective strategies to prevent errors. Each proposed strategy must then be validated based on criteria such as extra time taken to execute the task and the error rate with and without the strategies incorporated to the human interface component.

As a result of this work the authors intend to incorporate the process for strategy conception into a method for user interface conception – the MCIE [15] that will allow the designers to account for the human error when conceiving the user interface for automated industrial systems.

In the first session, this paper will present a review on human error classification. Then, in session two, it will give a brief description of the case study context, i.e. the operation of electrical systems. Section three presents the proposed process whereas session 4 gives an example on how to conceive strategies using the process. The paper concludes presenting the preliminary results of a validating experiment using a simulated environment and presents the future steps for this research.

2 The Human Error and Its Classification

A system is considered adequate when it behaves according to its design specification. A human action that modifies this behavior can be considered an error. According to Van Elslande and Alberton [13], every error can be considered human since it would have originated either from the human design or during human operation. Rasmussen [9], on the other hand argues that an error must not be considered an isolated result of a human action but rather the result of the interaction between the human operator and the system. Avizienis et al. [1] distinguish the result of an interaction and the interaction itself, which take the system into an invalid state, by categorizing them as an error and a fault respectively. To Laprie [7], the error can originate in the system (fault) or in the human operator (error); and a system fault can lead into a human error.

According to Holnagell [5] errors can cause incidents and accidents. An accident is defined as a short and sudden event that results in an unwanted situation directly or indirectly linked to a human action; whereas an incident is an unwanted event that ends a normal activity. Along this text the two concepts will be employed indistinctly.

Regardless of their differences most authors agree that the error originates during the interaction between the human agent and the system, and thus cannot be analyzed in an isolated manner neither from the system nor from the human points of view. Therefore, this is the viewpoint adopted in this work. It is also considered that the human error consists in an action that takes the system into an invalid state according to a predefined standard.

2.1 A Brief Review on the Human Error Classification

The human error has been classified according to different aspects. According to Reason's classification [10] it is based upon the user level of experience. His classification consists of: lapses that are made by experienced users who know the task and the work; whereas errors of intention happen when inexperienced users, due to lack of training, mistake the actions.

Swain's [12] classification is based on the task level execution mode. According to him there are: errors of omission, when parts of the task are omitted; errors of execution, when the task is executed in an incorrect manner; derivation errors, when an extra part is added to the task; errors of sequence, when the task sequence is altered and errors of timing, when the task execution time is altered.

Norman's [8] error classification is essentially based on the task action` level. He identified and classified errors according to the following patterns: (a) different tasks with initial actions in common and an unusual follow-up sequence; (b) a correct action performed on the wrong object; (c) sensory data interfere and unconsciously modify the course of action; (d) an internal association between thoughts and ideas; (e) loss of objective before concluding an action; (f) different devices have operation modes in common but with different meanings.

To conclude this brief review on error classification, Rouse & Rouse, apud Cellier [2], present an error classification based on Rasmussen's [9] cognitive model of task execution. This classification is effective to point out potential faults along the problem solving phases. Since this work intends to propose strategies for error prevention according to those phases. There is a particular interest in the execution phase that is characterized by Rouse & Rouse as follows: (a) omitting parts (actions) of a task; (b) executing a task repeatedly; (c) introducing a non prescribed action in the task; (d) executing actions out of a prescribed sequence; (e) inappropriate action timing; (f) incorrectly placed action; (g) task completed, but incorrectly and (h) task finished, but without completion (task goal was not achieved).

Analyzing the given classifications it becomes evident that many of the cited authors center their error analysis in the task execution phase and do not consider the preceding phases when cognitive factors lead the user into the error. To comprehend the error mechanisms it is necessary to extend the error analysis in order to clarify the mental processes behind the error and then propose effective strategies to prevent it. Therefore this paper proposes to extend the conventional error analysis which is based upon the history of errors in order to account for the characteristics of: the operator, the task and the work context.

3 Context Analysis

To better understand the error mechanisms it is necessary to understand the context in which it happens. According to Dekker [3], when analyzing an error incident, better than understanding the causes of the error is to understand why the information at the time of the error made sense to the system operator and why the specific reaction to it.

In automated systems the operator is a key figure during exceptional situations when the complexity of the required tasks requires human intervention. In spite of the importance of the human intervention, automation system designers tend to prioritize the system functionalities in detriment of the physical and cognitive abilities of their users.

This work is based on a study of human errors in the operation of electrical systems [4]. The authors had access to a corpus of study consisting of ten years of reports on human error that triggered accidents and incidents in the operation context of substations. The operation consists of performing maneuvers in order to put the substation in predefined configurations. These operations can be fully or partially automated, or manually executed by the human operators. Currently in the electricity industry, various levels of interaction with the plant are simultaneously available to the operator. It is possible to interact directly with the equipment panel; or through panels located in control rooms in the plant, as illustrated in Figure 1, and through supervisory system which represent the entire system allowing for the interaction through computer screens. During the interaction with supervisory systems, the information volume and content, presented to the operator is typically very high, allowing the access to a variety of devices and equipment statuses. Part of this information is available just for monitoring purposes whereas other demands acting within strict or hard deadlines. From the task point of view, the supervisory systems offer a completely new form of interacting with the plant. Figure 2 illustrates the screen of a supervisory system installed in the same electrical substation representing the plant information.

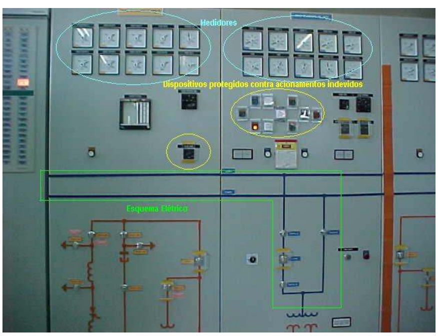


Fig. 1. Control pannel

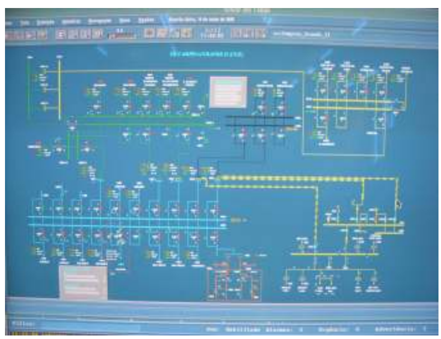


Fig. 2. Supervisory screen

4 Error Prevention Strategies - A Conception Process

To be effective, a strategy must anticipate potential errors looking into previously reported errors. To complement this approach one must also consider context factors that might lead into new error situations. Therefore it is proposed to analyze existing error reports using Rasmussen's [9] cognitive model of task resolution, adapted to the error classification proposed by Cellier [2]. It is also proposed to analyze other factors which might lead into novel error situations in the work context.

The process here proposed is based on the history of human triggered incidents and accidents in specific contexts. It considers the user profile, the work context and the task description. The process steps are illustrated in Figure 3.

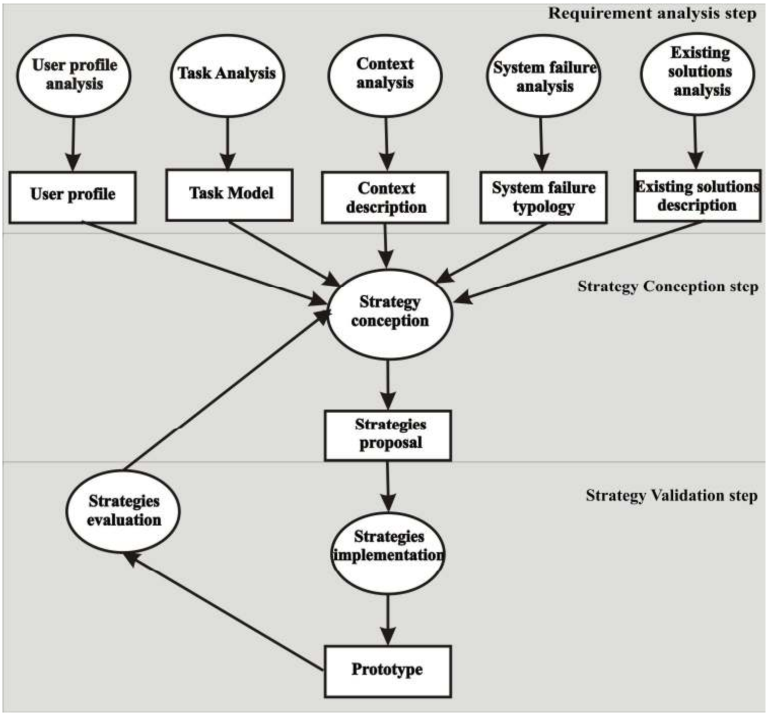


Fig. 3. Process to conceive strategies for error prevention

Knowing the user profile: characteristics and limitations; is essential to propose effective error prevention strategies. The proposed process is based on information collected in the context for which the strategies are proposed. The process does not propose an order for the information gathering, the order presented on this paper is just a suggestion, as represented in Figure 3.

One step in the process consists of raising the user profile. This is done by means of questionnaires answered by the system's operators. The questions cover professional aspects such as the level of expertise as well as personal characteristics such as age group, gender, etc.

Another relevant aspect is the task description, which includes information on: how often the task must be executed, how critical are the tasks results, how complex are the rules employed in the task execution procedure. This information can be gathered from interviews, observation of the work routine and documentation analysis.

The context data analysis is based on the norm ISO 9241 - part 11[6] and consists of the following steps.

- Data gathering: analyzes the physical, technical and organizational aspects of the work environment through documents. Build a questionnaire and a checklist, to gather more detailed information about specific ergonomic aspects, which were not clarified in the previous step. Apply the questionnaire to operators and use the checklist during visits to the work environment, as a guide to interviews and further observation.
- Data analysis: highlights the work context aspects that need to be controlled in order to prevent the human error.

Before proposing strategies it is important to review and critically analyze the effectiveness of the strategies adopted in the company. As part of the data gathering, one must collect information on previous strategies and their efficacy. This can be a source of invaluable information.

The strategies must be formulated in a clear manner, stating the kind of error to be prevented, how to be implemented and validated in the real environment. Before putting these in practice they must be validated. The validation can go through various levels that can consist of interviews, technical visits, and document analysis; but must be concluded with a pilot test in the real work context. A negative outcome must take the strategist back to step one of context analysis, i.e. data gathering. The proposed validation criteria should include the strategy influence on: the task duration, error incidence, and impact on the work routine, implementation viability and the impact on the operator's learning curve.

5 Applying the Process to a Substation Operation: A Case Study

The process was applied, as a case study, to the operation of an industrial an electric substation. This choice of context was made on the basis that this electric system is a critical one, where the human error can cause material losses as well as endanger the lives of the ones directly and indirectly involved with the system. This kind of task poses a high stress load on the operators.

The chosen substation is a real installation that belongs to one of the most important suppliers of the Brazilian electricity grid. The substation itself is an important node of the grid. There, the operators must be able to interact in any of the three levels described previously. The company, CHESF (Companhia Hidro Elétrica do São Francisco), has made available in a previous study a set of reports on human errors triggered accidents and incidents in a ten years period [4]. This set of reports constituted the main corpus of study for the process. It follows the description of each step in the strategy proposal process, applied to this case study.

5.1 The Operator Profile

To collect the data relative to the substation operator's profile (age, gender, background, levels of training and experience) it was applied the questionnaire Webquest [15]. The gathered profile describes an operator predominantly male, aged between 35 and 65, with the technical training acquired in the company; familiar with the use of computers and who worked mainly for the same company and in the same substation.

Therefore we are talking about an expert. The operators work environment is the substation control room and the patio where the equipment is placed.

5.2 Task Analysis

From technical visits to the installation, from observations, interviews and the analysis of the error reports, one arrived to the following task description. The task in the substation consists essentially in supervising and controlling the equipment to maintain or put the installation in previously defined configurations. To perform this task the operators interact with the supervisory system, the control panels in the control room and directly with the equipment panels in the patio. During routine or emergency situations, the task execution is formally bound by documents that prescribe the system final configuration and the procedure to reach it. Although, with time some of the procedures can be learned, the operators must strictly follow the documents. With time the procedures can change due to equipment replacement and installation upgrading.

From the error reports it was extracted the information that routine tasks, particularly the more frequent ones, are strongly related to errors. On the other hand, the error incidence during rare programmed tasks is higher than during urgency and emergency procedures. Simple tasks and emergency tasks had the highest error incidence.

5.3 The Work Context

The information on the work context was gathered through guided interviews, and observations based on checklists built according to the standard ISO 9241, part 11. The intention was to investigate and highlight ergonomic aspects of the operator's task. From the analysis it was found that the operators work in pairs and in shifts. It was also registered complaints of work fatigue and cognitive overload. The error reports registered situations when steps of the prescribed procedure were bypassed during routine situations. From the equipment point of view, it was found that equipments with similar functions but from different manufactures were not compatible in terms of interaction rules, leading into errors. From the organizational point of view the operators were generally unsatisfied with the management of faulty behavior. From the physical point of view there was dissatisfaction with the acoustics and lighting in the control room.

5.4 Error Analysis

The error report analysis step was based on the Rasmussen's [9] cognitive model of task resolution, adapted to the error classification proposed by Cellier [2]. The corpus analyzed consisted of 35 human error reports related to incidents and accidents in the company spread over a period of 10 years. It follows the findings according to each phase of the task resolution Rasmussen's model.

System observation: in 32 out of the 35 cases, the error happened during this phase. *No observation* and *incorrect observation* responded for 62% of the reported causes of error. Hypothesis proposal: in 11 reports out of the 35 cases no hypothesis was formulated. From the analysis it was found that not formulating a hypothesis and formulating either an insufficient or inconsistent one responded for 56% of the problems during this phase. Hypothesis evaluation: during this phase, 86% of the errors were associated to accepting the hypothesis or not evaluating one. Goal setting: From the study it was found that in 72% of the cases the goal set was correct, in the remaining 28% cases were either incorrect or incomplete. Choice of procedure: 17 out of the 35 reports mentioned the correct choice of procedure whereas 12 reports mentioned incorrect choice and 6 mentioned the choice of an incomplete procedure. Task execution; this phase evidenced a high incidence of cases where the procedures were not completed (6 out of 35) and where the intermediate actions were either performed on the wrong device (15 out of 35) or were omitted (3). The remaining cases accounted for actions out of a sequence or inappropriate timing. An example of performing the correct action over the wrong device in this context is when the operator follows the prescribed procedure and closes a wrong switch break.

5.5 Previous Strategies

During the technical visits to the substation installation it became evident the efforts made by the technical personnel to prevent errors during the operation. It follows the description of some of these initiatives.

Placing a physical barrier over the interaction devices (switches, buttons, etc.) located on control panels, in order to delay the action and give the operator a chance to reflect over it. Another strategy consists in demanding the operator to confirm with his colleague each step (action) in the procedure and to acknowledge the system feedback to this action. It has also been placed a safety warning in the form of a yellow strip painted around the equipment to avoid unintended actions on the panels (accidently pressing a button) due to close proximity. Restriction notes, in the format of warning cards, are also placed by the side of an interaction device on the equipment panels, to warn the operators of a restriction related either to safety or to temporary unavailability of equipment due to maintenance.

5.6 Strategy Proposal

The following strategies were proposed on the basis of: relevance (the error report studies highlight the cognitive phases of task execution when the operator is more likely to err); operator profile (skills and limitations); task description (relevance and impact of not being completed successfully); knowledge of the context (work conditions) and the effectiveness of the error prevention strategies already in use. The proposed strategies are the following. To increase the visual distance between similar interaction devices found on the panels (buttons, switches, etc.). To review the use of terms in the panels user interface to ensure a standard. Visually highlight the objects involved in the programmed procedures. Provide a mechanism to indicate the sequence of actions to be performed during the execution of a procedure, highlighting the action currently being executed in relation to the sequence to be followed (evolution along the sequence). Generate an alert when an action being performed is missing or unduly introduced in the sequence. Call the operator's attention to time restriction during the procedure execution. Block an action which is out of a sequence or overdue, informing the operator of such prohibition.

5.7 Validating the Strategies

Given the practical impossibility of validating the proposed strategies directly in the real work environment these have been tested in a simulated work environment. This environment consists of a virtual reality representation of the control room offering the two abstraction levels of interaction: control panels and the supervisory system work station. The simulator [11] represents the plant behavior using Petri net models. The two levels are interconnected and the effect of any action performed in one level will cause an update in the other.

The strategies were initially tested using the simulated environment, nonetheless these are still to be validated in the real work environment, in a pilot situation, adopting the criteria mentioned before. The validation protocol is currently under discussion with the company. The preliminary results of the pilot testing in the virtual work environment already indicate a positive influence on the error rate.

6 Concluding Remarks

Given the extent of the consequences of the human error during the operation of industrial systems, a process for error prevention strategies can be of great help to planners and designers. Usually, error prevention strategies are conceived on the basis of specific events and rarely consider the human interface component. The knowledge of the task, the operator profile, and the error episodes issued by the process has resulted in effective strategies to prevent the error; since it can anticipate potential errors and not just prepare for already known situations. The following steps in this research will consist in validating the strategies in the real work context as well as applying the process to a different work context. The objective is to investigate how effective the strategies can be in reducing the operator's cognitive load, the human error rate, thus improving operation safety.

References

1. Laprie, A.J.C., Randell, B., Landwehr, C.: Basic Concepts and Taxonomy of Dependable and Secure Computing. IEEE Trans. on Dependable and Secure Computing 1(1), 11–33 (2004)
2. Cellier, J.M.: L'erreur humaine dans le travail. In: Leplat, Jaques, De Terssac, Gilbert. Les facteurs humains de la fiabilité dans les systemes complexes, Avec la colaboration de: J.M. Cellier, M. Neboit, A. Oudiz. Octares, Marseille (1990), ISBN 2- 906-769-03-7
3. Dekker, S.: The Field Guide To Human Error Investigations. Ashgate Publishing Co. (2002)
4. Guerrero, C.V.S., Vieira, M.F.Q., Mercantini, J.M., Santoni, C.: A process for Human Centered Modelling of Incident Scenarios. In: Holzinger, A. (ed.) USAB 2008, 1st edn. LNCS, vol. 5298, pp. 439–458. Springer, Heidelberg (2008)
5. Hollnagel, E.: Barriers and Accidents Prevention. In: ASHGATE Pub Ltd, ed. (2004), ISBN: 0754643018
6. ISO 9241-11. Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability. International Organization for Standardization, Geneva, Switzerland (1998)

7. Laprie, J.-C.: Guide de la sûreté de fonctionnement. Cépaduès - Editions., Toulouse, p. 324 (1995)
8. Norman, D.A.: Categorization of action slips. Psychological Review (1981)
9. Rasmussen, J.: The Definition of Human Error and Taxonomy for Techincal System Design. John Wiley & Sons Ltd, Chichester (1987)
10. Reason, J.: Human Error. Cambridge University Press, Cambridge (2006)
11. Santoni, C., Turnell, M.F.Q.V., Mercantini, J.M., Scaico, A., Nascimento, J.A.N.: A real-time interface simulator for operator's training: a proposed architecture. In: Summer Simulation Multiconference, 2007, San Diego, vol. 1, pp. 460–467. The Society for Modeling and Simulation International (2007)
12. Swain, A.D.: Human factor associated with prescribed action links. USA Sandia Laboratories, Sand (1976)
13. Van Elslande, P., Alberton: L'Accident de la Route: Chercher l' Erreur. Securité et Cognition, Hermes Editions (1999)
14. Vieira, M.F.Q.: Accounting for Human Errors in a Method for the Conception of User Interfaces. In: International Mediterranean Modeling MultiConference - I3M 2004, Genoa, Italy, Proceedings of I3M 2004, Bergeggi Italia, vol. 1, pp. 122–130 (2004)
15. Vieira, M.F.Q., De Queiroz, J.E.R., de Oliveira, R.C.L.: WebQuest: a configurable web tool to prospect the user profile and user subjective satisfaction. In: 11th International Conference on Human-Computer Interaction, HCII 2005. MIRA Digital Publishing, Las Vegas (2005)

Wearable Modular Device for Facilitation of Napping and Optimization of Post-nap Performance*

Djordje Popovic[1,2], Giby Raphael[1,2], Robin Johnson[1], Gene Davis[1], and Chris Berka[1]

[1] Advanced Brain Monitoring, 2237 Faraday Ave., Suite 100, Carlsbad, CA 92008
[2] University of Southern California, Los Angeles, CA 90089
dpopovic@b-alert.com

Abstract. Sleep deprivation-induced deficiencies in performance can be associated with high financial and human costs. Napping is an effective countermeasure, but the effects depend on previously accumulated sleep debt and timing, duration and sleep architecture of the naps. Long-term assessment of sleep architecture of nap/sleep episodes could yield an estimate of the accumulated sleep debt and help optimize the napping schedule. Moreover, sensory stimulation coupled with real-time assessment of sleep states could optimize sleep architecture and duration of each nap. With these goals in mind we designed a wearable device, dubbed Nap Cap, which integrates real-time EEG analysis with audio, visual and thermal stimulation. The prototype was evaluated on seven subjects (fully rested vs. sleep-deprived). While the prototype provided high quality EEG and comfort, sensory stimulation did not significantly influence sleep architecture. Evaluation of more paradigms of sensory stimulation on larger samples is warranted before final conclusions can be made.

Keywords: Nap, Sleep Deprivation, Performance Optimization, Wearable Devices.

1 Introduction

Sleep restriction has a profound impact on human behavior, performance and physical health. Even small amounts of sleep loss accumulate over time resulting in a “sleep debt”, and manifest in impairments of alertness, memory and other cognitive functions [1]. Deficient performance can be associated with significant social, financial and human costs. Impaired vigilance is the leading cause of transportation and industrial accidents in the US [2, 3], while recent National Aeronautics and Space Administration (NASA) reports revealed that pilots often experience brief episodes of sleep while flying [4]. Due to the large number of shift workers or workers with irregular schedule sleep restriction is considered a serious public health and safety concern [5].

While temporary amelioration of the effects of sleep restriction on motor and cognitive performance can be achieved by natural or pharmacological stimulants (e.g. caffeine) or overt stimulation of the senses, brief naps taken at appropriate times throughout the day are the only intervention without long-term adverse effects that

---

* Approved by DARPA for public release. Distribution unlimited.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 289–298, 2009.

efficiently counter chronic sleep restriction. Unfortunately, attempts at designing optimal napping schedules [6] have been unsuccessful. The main reason for failure was the narrow focus on duration and timing of naps, whereas the influence of other factors such as sleep architecture, sleep debt accumulated prior to a nap and the subject's susceptibility to sleep deprivation, was neglected. However, adequate amounts of stable NREM Stage 2 sleep in each nap [7] and long-term balance among NREM, REM and slow-wave sleep [1, 6] have been identified as key determinants of post-nap performance. These findings suggested that manipulating the proportions of the key sleep stages during napping could substantially reduce the number and/or duration of naps necessary to optimize motor and cognitive performance.

Our group has recently developed a modular device for optimization of napping in operational environments. Dubbed the "Nap Cap", the device is designed to assess sleep architecture of each nap in real-time by measuring brain electrical activity (EEG), maintain a record of all naps taken, provide protection from environmental disturbances, deliver sensory stimulation to influence the sleep architecture in the desired way and awaken the subject at an appropriate time to avoid sleep inertia (Fig.1A). This article presents the results of a pilot evaluation of a prototype device.

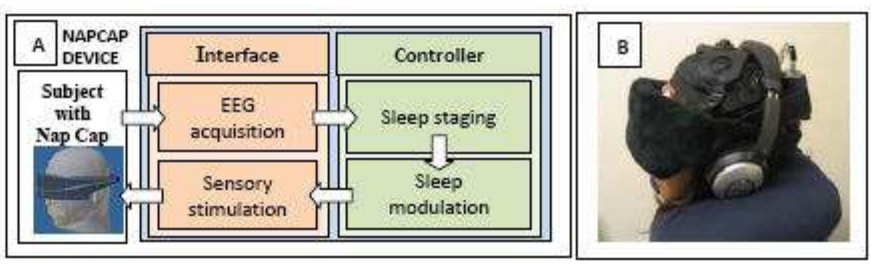


Fig. 1. A. Architecture of the Nap Cap device consisting of four main modules. B. Prototype device assembled of off-the-shelf component.

2 Methods

2.1 Prototype Nap Cap Device

The prototype Nap Cap device (Fig.1B) was assembled with off-the-shelf components. EEG was recorded with a wireless wearable 9-channel headset (Advanced Brain Monitoring Inc., Carlsbad, CA). Recorded channels included C3-A2, C4-A1, Fp-Fp2, Fz-PO, Cz-PO, vertical and horizontal electrooculogram and submental EMG. Bose® noise cancelling headphones and an eye cover provided protection from environmental noise and light respectively. The headphones also delivered auditory stimulation, whereas two blue LED arrays built into the foam of the eye cover provided visual stimulation. Blue light was chosen because of its reported beneficial effect on duration and severity of post-nap sleep inertia [8]. An inflatable neck pillow with a battery powered heating element ensured increased comfort, and provided thermal stimulation of the neck which was expected to produce effects similar to the well known 'hot bath effect' [9].

2.2 Study Design

Seven healthy subjects (three females, age range: 22-25 yrs) participated in the study. The subjects reported no significant previous or existing health problems, including substance abuse, and specifically, had no sleep-related complaints.

The subjects took a 1-hour midday nap wearing the Nap Cap on two separate occasions: fully rested, and after sleep deprivation. The rested and sleep-deprived sessions were between one and seven days apart. The subjects completed sleep logs and wore actigraphs for three days prior to either session so that their sleep schedule could be assessed. Subjects were required to abstain from caffeinated beverages on the days the experimental sessions were conducted.

On the night before their second session the subjects were allowed only 2 hours of sleep. Compliance with the instructions was enforced by requiring that the subjects leave phone messages and send emails every 30 minutes between 12AM and 6AM except for the 2 hours they were allowed to sleep. The compliance was further confirmed by a visual inspection of the actigraphic data for all seven subjects.

2.3 Experimental Sessions

Both experimental sessions consisted of four main parts (Fig.2). The subjects were evaluated with the proprietary battery of psycho-physiological tests called Attention and Memory Profiler (AMP) before and after the nap during each session. Upon completion of the pre-nap AMP, the subjects were given lunch and then took a 60-minute nap wearing the Nap Cap prototype. The subjects completed a brief questionnaire about comfort of the Nap Cap immediately after the nap, and then started the post-nap AMP evaluation.

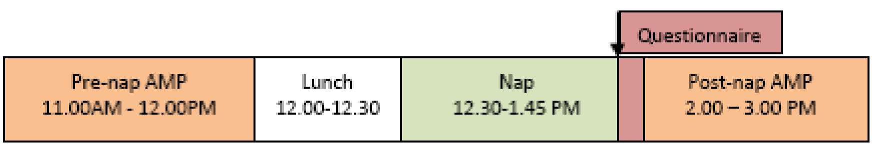


Fig. 2. Experimental protocol on both experimental sessions

AMP evaluation: The AMP evaluation included three tests of vigilance and memory. The 3-Choice Vigilance Test (3C-VT) required the subjects to discriminate primary (70%) from two secondary geometric shapes and respond as quickly as possible over a 20-minute test period. The memory tests, each 7 minutes long, were variants of the Image Recognition Test (IR) where the subject had to memorize 20 images and identify the 20 training images among 80 previously unseen testing images. In Numbers Paired Associate Learning Test, (N-PAL), a number was assigned to each image and subjects had to identify the correct image-number pairs. Five-minute breaks were given to the subjects in between the tests, thus the total duration of the AMP evaluation was little less than an hour.

Napping: The subjects napped in a room at the temperature of 24-25°C, lying supine in a comfortable chair with the back tilted so that the angle to the ground was

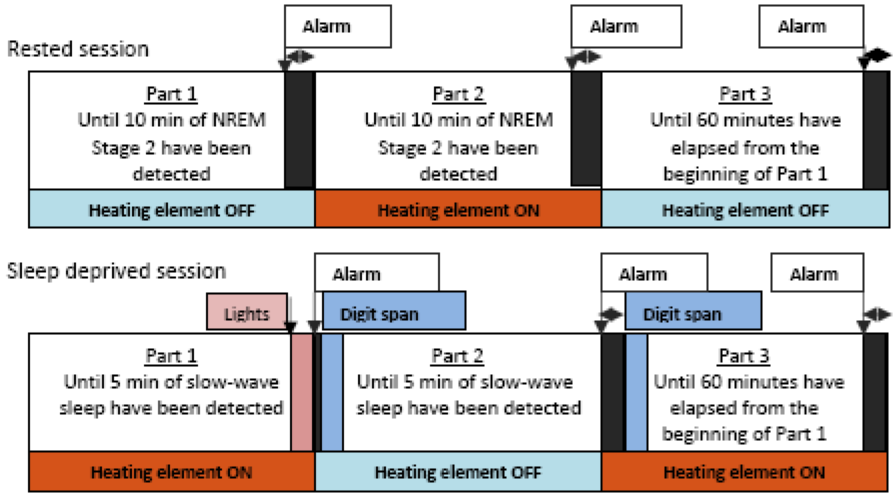


Fig. 3. Experimental manipulations during rested (above) and sleep-deprived session (below)

approximately 30^{o}. The lights in the room were turned on in order to test the efficacy of the eye cover in blocking the environmental light. The subjects were constantly monitored through a camera mounted next to the chair, and the recorded EEG was transmitted in real-time to the computer in front of the experimenter.

On both sessions napping consisted of three parts during which the heating element was alternately turned on and off (Fig.3). Such design aimed at separating effects of increasing the local skin temperature on sleep variables (see below) from the expected effects of sleep deprivation and sleep propensity. The emphasis during the rested session was on interventions facilitating transitions from wakefulness to solid sleep. Therefore, the subjects were not allowed to progress beyond NREM Stage 2, and were awakened with a 2-minute long alarm after they had spent 10 minutes in stable sleep. In contrast, during the sleep deprived session the emphasis was on procedures facilitating a smooth transition from deep sleep to wakefulness with minimal post-nap sleep inertia. Consequently, the nap was interrupted after 5 minutes of slow-wave sleep with either light from the LED arrays, or the alarm sound delivered through the headphones. If slow-wave sleep had not occurred, the nap was interrupted after 30 minutes while the subject was in deep Stage 2 sleep. The LEDs were activated for 90 seconds only during the first interruption. In case the subjects did not wake up during this manipulation, the alarm sound was played briefly (30 seconds) to awaken them. On subsequent interruptions only a 2-minute alarm sound was used to awaken the subjects. Upon each interruption the subjects were asked to complete an auditory digit span test (2 sequences of 7 random digits that the subjects were supposed to reproduce immediately after hearing them). Failure to correctly reproduce the digits was interpreted as a presence of sleep inertia.

2.4 Outcome Measures

Sleep staging and sleep variables: The EEG records of the naps were scored by a board-certified sleep specialist in 30 second epochs according to the AASM standard rules [10]. Recorded EEG contained very few artifacts. Generated hypnograms were then used to calculate total sleep time (TST), sleep efficiency (SE), number of spontaneous awakenings after sleep onset (WASO), % time spent in each sleep stage, sleep onset latency (SOL), latencies to NREM stage 2 and SWS (Table 1 and 2, and Figure 4). Sleep onset latency (SOL) was defined as time from the beginning of each part of a nap till the first occurrence of 3 or more consecutive epochs of Stage 1 or 1 epoch of any deeper stage of sleep. Latencies to stage 2 and SWS were defined as the time from the sleep onset till the first epoch of stage 2 or slow-wave sleep respectively.

Performance measures: included reaction times and percentages of correct responses on AMP tests as well as the digit span test. Any error during reproduction of digits was interpreted as a presence of sleep inertia.

Questionnaire: Comfort of the Nap Cap prototype was evaluated with a questionnaire which the subjects filled after the nap on both their rested and sleep-deprived session. The subjects graded perceived comfort for each component (EEG Headset, pillow, eye cover, headphones) and the device as a whole on a nominal scale with 5 categories that were later converted into ranks for the purposes of analysis (1-very uncomfortable, 2-uncomfortable, 3-neutral, 4-comfortable, 5-very comfortable).

Statistical analyses: The effects of experimental conditions on the sleep variables were tested by a 3-way ANOVA, with session (rested, sleep-deprived), time (Part 1 or 2) and heat (off, on) as the factors. Factor Time essentially modeled sleep propensity – it was expected that latencies to all stages will be shorter and % time spent in deeper sleep stages (Stage 2 and SWS) bigger in Part 2 than Part 1. Only main effects were tested because of the small sample size which created numerical problems in a model with interactions. Part 3 was excluded from the analysis because its duration was very variable (from 5 to 15 minutes) during the rested session, while most subjects did not complete it at all on the sleep-deprived session. The results of the questionnaire were analyzed only descriptively.

AMP performance measures were analyzed with a 2-factor RMANOVA with Session (rested, sleep-deprived) and Condition (pre- and post-nap) as the factors.

3 Results

3.1 Effects of Sleep Deprivation, Sleep Propensity and Thermal Stimulation on Sleep Variables

Consistent with our expectations, latency to Stage 2 ($F_{(3,18)}$=3.8, p=0.05) and % time spent in Wake ($F_{(3,18)}$=6.05, p=0.023), and Stage 1 ($F_{(3,18)}$=7.57, p=0.013) decreased while TST ($F_{(3,18)}$=6.01, p=0.027), SE ($F_{(3,18)}$=6.45, p=0.017) and % time spent in SWS ($F_{(3,18)}$=4.22, p=0.054) increased after sleep deprivation. SOL was significantly shorter in Part 2 in both sessions ($F_{(3,18)}$ =4.85, p=0.041). The magnitudes of all the significant effects can be inferred from Figure 4 and Tables 2 and 3.

Contrary to our expectations, the heating did not seem to have had any effect on the sleep architecture. The only variable that was (marginally) affected by heat was the number of awakenings after sleep onset (3.8±2.2 without vs. 2.3±1.7 with heat, p=0.08).

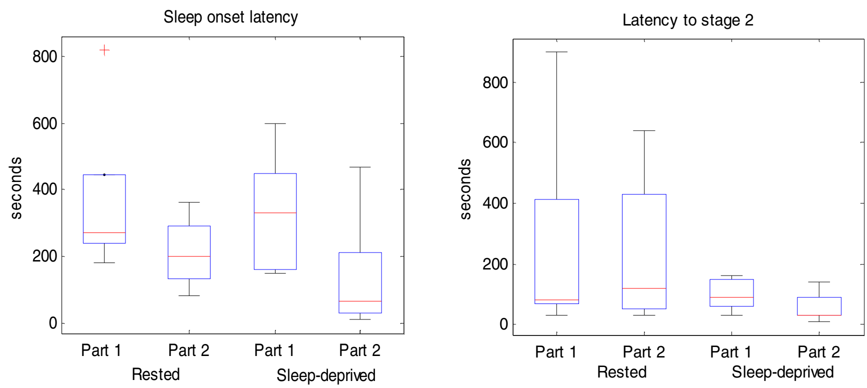


Fig. 4. Sleep onset latency and latency to Stage 2 (both in seconds) across various parts of the experiment. Red lines – medians; blue boxes - lower and upper quartile; black whiskers - full range.

Table 1. Sleep architecture of naps with subjects fully rested (SE in %, all other variables in minutes)

Subject	TST	SE	Wake	S1	S2	SWS
1626	52.8	84.4	14.1	18.6	65.8	0.0
1627	41.6	70.4	28.5	13.5	55.5	1.4
1628	31.8	64.9	33.7	17.4	47.6	0.0
1629	42.3	71.9	26.9	23.8	48.2	0.0
1630	39.8	65.5	33.4	46.0	19.5	0.0
1631	38.3	61.3	37.6	25.9	35.5	0.0
mean±sd	**41.8±8.2**	**69.7±8.1**	**29.0±8.3**	**24.2±11.6**	**45.3±16.1**	**0.2±0.6**

Performance measures: Reaction times and percentages of correct responses showed the expected pattern of performance deterioration following sleep deprivation, and its restoration after a nap (Fig.5). However, none of the visually apparent differences was statistically significant, probably due to the small sample size and high within-subject variability. No subject failed the digit span test.

Comfort: The subjects found the Nap Cap comfortable (grade: 3.3±0.7) with the EEG headset being the least comfortable component (grade: 2.7±0.6). The most frequent complaint/suggestion about the prototype as a whole was that it contained too much

Table 2. Sleep architecture of naps following sleep deprivation (SE in %, all other variables in minutes)

Subject	TST	SE	Wake	S1	S2	SWS
1626	45.3	76.0	24.0	15.0	50.8	10.0
1627	48.7	84.8	12.2	02.2	52.9	32.7
1628	47.5	78.7	21.0	13.8	64.9	00.0
1629	44.8	72.7	27.3	15.1	57.6	00.0
1630	47.0	82.0	18.0	12.2	36.6	33.4
1631	45.8	82.8	17.2	07.8	43.4	31.6
1632	51.5	85.6	14.4	21.8	63.7	00.0
mean±sd	**47.2±2.3**	**80.4±4.8**	**19.2±5.3**	**12.6±6.2**	**52.8±10.4**	**15.4±16.4**

stuff and should be lighter and less obtrusive. Grades were consistently (although insignificantly) higher on all items on the sleep deprived session, which may reflect a change in perceptual threshold due to sleep deprivation, or an adaptation to wearing the Nap Cap. The two subjects who gave the lowest grade for comfort had the lowest sleep efficiencies and highest % of stage 1. Their complaints had however little to do with the Nap Cap: one was uncomfortable about sleeping under surveillance and another complained of not being able to sleep well in the supine position.

4 Discussion and Conclusions

Contrary to common reports of low signal quality in unattended sleep recording with ambulatory polysomnographs the Nap Cap prototype delivered high quality EEG with very few artifacts, demonstrating that careful mechanical and electrical design can result in a robust yet easy to use wearable EEG acquisition system. Algorithms for real-time sleep staging have been developed and successfully tested on a larger data set [11]. However, the EEG headset used in the Nap Cap prototype acquires many more channels that it is needed for accurate sleep staging, and utilizes modified 'wet' Ag/AgCl electrodes with a small amount of conductive paste, which is a clear disadvantage for in-field use. The next generation of Nap Cap will be designed to use forehead and peri-orbital dry electrodes, possibly similar to those used in the ARES™ Unicorder. Forehead EEG in combination with head actigraphy suffices for accurate automated distinction among wakefulness, REM, light NREM and slow-wave sleep [12].

While the sleep monitoring in real time was successfully accomplished, sleep modulation by means of sensory stimulation achieved very little effect on the transition from wakefulness to sleep. Thermal stimulation of the neck that we hoped would have replicated the 'hot bath' effect bore no objective impact on sleep architecture of the naps, and some subjects even found it distracting. The heating was however not excessive since the skin temperature increased in all subjects for less than 4°C during

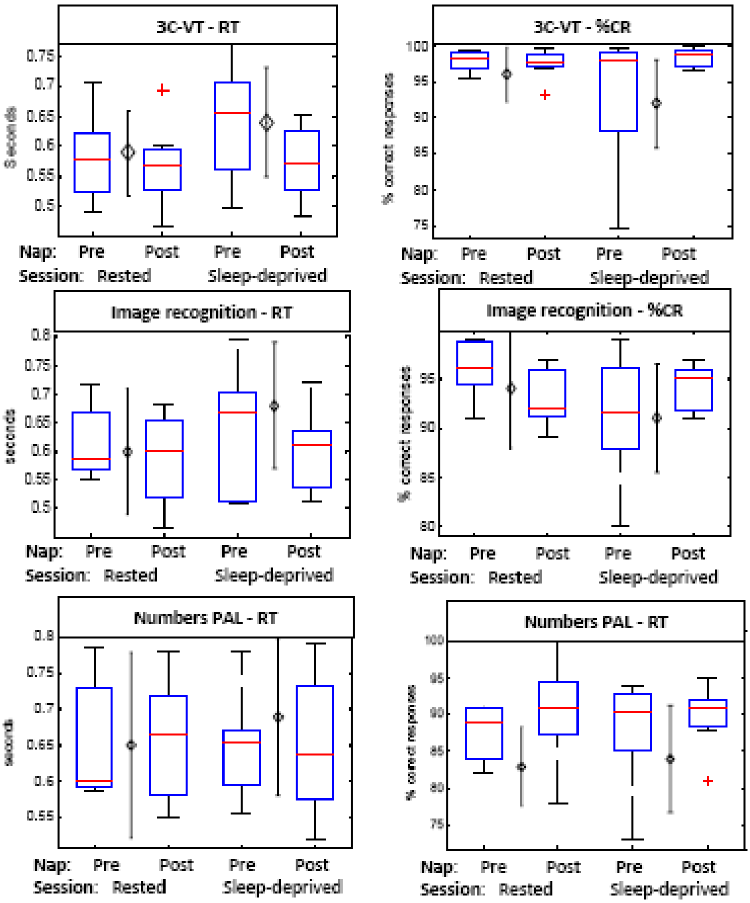


Fig. 5. Mean reaction times (RT) and percentages of correct responses (%CR) on AMP tests. Black error bars between the box plots represent the mean and standard deviation of the same variables from a reference database of over 1,000 rested and sleep-deprived subjects.

the parts of the experiment when the heating element was switched on (rested session: 2.18±1.01°C, sleep deprived session: 2.81±0.52°C, in both cases measured after 20 minutes). The measured increase in skin temperature is comparable to that in recent reports of the successful use of thermal suits to increase the depth and quality of sleep [9]. The failure of thermal stimulation in our experiment could perhaps be explained by the substantially smaller area of the skin that was heated with the Nap Cap. Efficient heating of larger areas of the skin would be difficult to achieve with a battery powered system without compromising the lifetime of usage before the batteries need to be replaced or recharged. We plan to conduct more studies of the effects of thermal stimulation on sleep architecture, and build eventual positive findings into an improved version of the Nap Cap.

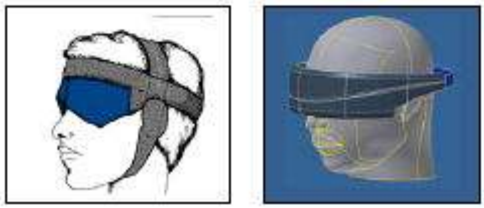


Fig. 6. Two potential form factors: a soft ski-mask design (left) and a more rigid visor design (right). In both cases the electronics and batteries will be housed in the appropriately designed eye cover.

Awakened by the applied auditory stimulation the subjects showed no gross signs of residual sleep inertia on the digit span test. The applied visual stimulation on the other hand did not awaken most of the subjects, and there was no evidence that it added anything to the subsequent brief auditory stimulation in terms of mitigating the post-nap sleep inertia. It is however unclear whether subtle deficits in motor or cognitive performance would have been noted, and differences between the visual plus auditory vs. auditory stimulation alone detected if a more sensitive (but still brief) test had been applied upon each awakening. Furthermore, it would be premature to draw general conclusions on the basis of few fixed stimulation patterns that have been tested so far. Our future research will focus on designing a variety of the stimulation patterns, exploring more sensitive metrics for quick detection of residual sleep inertia, and improving the experimental design of the future validation studies. Comfort of the Nap Cap also needs to be further improved as it is crucial for success of a device that aims at optimizing sleep. This will be achieved by reduction in the number of sensors and weight of electronic circuitry, custom design of the earphones, and integration of all components into one-piece easy-to-apply device (figure 6).

In conclusion, the results of the pilot validation confirmed the viability of the concept of using an EEG-based wearable device to monitor and optimize sleep, but more psycho-physiological research and further technical improvements are warranted before it becomes ready for field use.

Acknowledgements

DARPA SBIR grant W31P4Q-08-C-0123, NIH SBIR R44HL068463-05 and 2R44-DE016772-0

References

1. Banks, S., Dinges, D.F.: Behavioral and physiological consequences of sleep restriction. Journal of Clinical Sleep Medicine 3(5), 519–528 (2007)
2. Federal Highway Administration Office of Motor Carriers (1996), Truck and bus accident fact book, Washington, D.C. (1994–1996)
3. Dinges, D.F.: An overview of sleepiness and accidents. Journal of Sleep Research 4(S2), 4–14 (1995)

4. Rosekind, M.R., Gregory, K.B., Miller, D.L.: Crew Factors in Flight Operations XIII: A Survey of Fatigue Factors in Corporate/Executive Aviation Operations. NASA Ames Research Center, Moffett Field (2000)
5. Mitler, M.M., Carskadon, M.A., Czeisler, C.A., et al.: Catastrophes, sleep, and public policy: consensus report. Sleep 11(1), 100–109 (1988)
6. Takahashi, M.: The role of prescribed napping in sleep medicine. Sleep Medicine Review 7(3), 227–235 (2003)
7. Hayashi, M., Motoyoshi, N., Hori, T.: Recuperative power of a short daytime nap with or without stage 2 sleep. Sleep 28(7), 829–836 (2005)
8. Munch, M., Kobialka, S., Steiner, S., et al.: Wavelength-dependent effects of light exposure on sleep architecture and EEG power density in men. American Journal of Physiology 290(5), R1421–R1428 (2006)
9. Rayman, R.J., Swaab, D.F., Van Someren, E.J.: Skin deep: enhanced sleep depth by cutaneous temperature manipulation. Brain 131, 500–513 (2008)
10. American Academy of Sleep Medicine. The AASM Manual for the scoring of sleep and associated events. AASM, Westchester, IL (2007)
11. Popovic, D., Johnson, R.J., Westbrook, P., et al.: Automated sleep staging in real time using a single forehead (Fp1-Fp2) EEG channel. In: APSS 2009 (Submitted, 2009)
12. Popovic, D., Ayappa, I., Hauri, P., et al.: Accuracy of automated sleep staging using signals from a single forehead site. Sleep. Abstract Supplement 31, A332 (2008)

Converging Minds: Assessing Team Performance Using Psychophysiological Measures

Aniket A. Vartak, Siddharth S. Somvanshi, Cali M. Fidopiastis, and Denise Nicholson

Institute for Simulation and Training
3100 Technology Parkway
Orlando, FL, 32826 USA
{avartak,ssomvans,cfidopia,dnichols}@ist.ucf.edu

Abstract. Effective teams are an integral component to the success and the advancement of any organization. This issue emphasizes the need to develop valid measures for team performance especially in operational environments. The use of psychophysiological data has been proposed as a candidate for developing these team-level measures. In this paper, we review past research in the field and discuss two contrasting approaches to model human cognition used in the context of teams. We then propose a test-bed for evaluating these models for human-in-the loop adaptive systems using psychophysiological measures.

Keywords: Team Performance, Team Cognition, Psychophysiology, Social Cybernetics, Information Processing, Closed-Loop Human Systems.

1 Introduction

Improvements in team performance are related to team members' understanding of the shared mental model (SMM) represented within the team [1]. This understanding implies that each team member knows his/her own capabilities, the task at hand, and the capabilities of the other teammates. Good team members use this information to mentally simulate how others on the team will react in different situations [2]. More specifically, SMM consists of the following factors: team cognition, team skills, team attitudes, team dynamics and team environment [3]. Team cognition is defined as a mechanism that produces coordinated behavior, emerging from the interplay between each team member's individual cognition and team process behaviors such as coordination and communication [4]. Understanding team cognition is a key aspect for predicting team performance [1].

Measurement of team cognition is still in its initial stages. The lack of research in this field [5] may be attributed to the inadequate development of the construct itself along with confusion over how these cognitive variables can be measured at a team level. Even so, theories governing the construct of team cognition continue to be solidified with the establishment of related terminology and methodologies [4]. Studies in several laboratory settings have provided a better understanding of the candidate techniques for measuring team cognition. In this paper, we will describe relevant results from such studies. We also describe two opposing views on cognition within

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 299–303, 2009.

individual/teams, namely social cybernetics and information processing. Finally, we propose a test-bed for evaluating these two models using psychophysiological measures.

2 Previous Work

Kiekel et al. [6] attempted to use voice communication data to evaluate team performance. In this study, the authors collected communication logs from a team of three members performing a task of flying a simulated plane over 10 missions. Changes in dominance patterns (how much each team member spoke) for each mission were then analyzed. The results showed that higher numbers of distinct dominance patterns in a mission correlated with poorer team performance.

Another study by Henning et al. [7] made use of psychophysiological data to determine team performance. The authors applied their cybernetic model of social-psychophysiological compliance (SPC) and evaluated it as a predictor of team performance. SPC, in this regard, predicts that psychophysiological measures between team members will synchronize when team performance is optimal because of the ability of good team members to anticipate each other's responding. As a part of this study, 18 teams of two participants each were tasked with manipulating a simulated object through a complex two dimensional path. SPC was calculated from heart rate variability (HRV), skin conductance response (SCR) and respiration data and cross correlated between team members. The results showed significant coherence among the psychophysiological measures for high performing teams. Based on these results, the authors claim SPC not only effectively predicts team performance, but provides a reliable means to trigger adaptive automation.

3 Social Cybernetics

The study by Henning et al. [7] is based on the cybernetics perspective. This perspective views motor behavior as a means of self-regulation via effects of motor activity on cardio-respiration, hormonal activity and other physiological systems in addition to its role in body locomotion. [8]. The cybernetics perspective for one person is extended to a social context with multiple persons interacting with each other. This theory is based upon the hypothesis that an individual can control sensory feedback not only from their own behavioral movements, but also from others with whom he/she is interacting. The cybernetics approach applied to teams is in contrast to the information processing approach, which views all motor activity as end event following series of mental processing steps. In an extension of the virtual object manipulation study described above Henning et al. [9] evaluated the use of SPC as a predictor by varying the difficulty level of task as a function of the SPC metric. In *matched condition* the difficulty level was increased when SPC indicated that the team could handle increased task demand and lowered when SPC indicated that the team could not. In the *unmatched condition,* the difficulty level was decreased when SPC indicated that the team could handle increased task demand and vice versa. Task performance was analyzed for both conditions and the error of tracking was found to be lower in the *matched condition.* However this was accompanied with an increase in task completion time.

4 Information Processing

The social cybernetics view opposes the information processing view, in which the motor actions performed are considered as the end result of mental process. Within an augmented cognition framework, an information processing approach focuses on determining the instantaneous cognitive load using physiological sensors. Mitigation strategies such as task scheduling, modality encoding are sketched to cope for individuals performing under stress. Task sharing and offloading is mentioned as an addition when the approach is extended to a team environment [12]. Other mitigation strategies such as automation of information acquisition and automation of information analysis are also suggested, with a caution that automating the decision making process would hinder the team performance [13]. The effectiveness of these mitigations is usually measured with the NASA Task Load Index [14]. The use of psycho-physiological sensors to predict the mental workload is an unobtrusive method with the potential to estimate the workload in real time.

5 Test-Bed for Evaluation of Different Theories

The challenges for bringing the measurement into an operational environment are numerous. The architecture described in [10] was developed to provide mitigation to a single operator with psychophysiological sensors. In this architecture the cognitive state of the operator is estimated by multiple sensor data streams. Salient features extracted from the streams are further classified into levels of cognitive states. This architecture can be extended to a team of operators as shown in Figure 1. Using the social cybernetics model described above, the sensor data from a team of operators could be combined into a set of features (such as metric of similarity of HRV, GSR data) and these features could be classified into levels of compliances. These levels could then drive a mitigation which adapts to the over-all level of compliance between the team members to suggest an appropriate strategy.

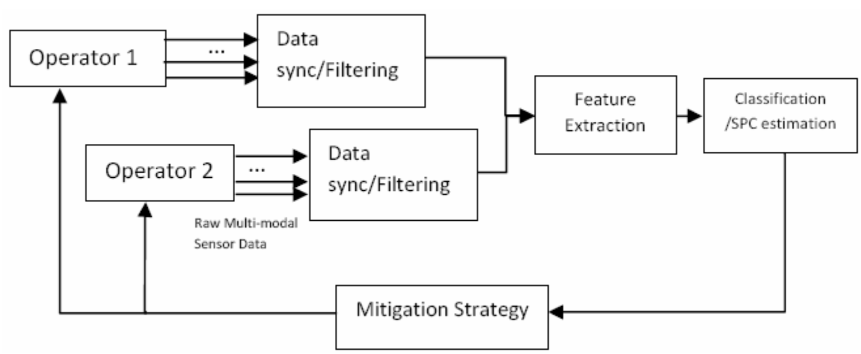


Fig. 1. Test-bed based on social cybernetics model

To test the information processing model for teams, cognitive state estimation of each of the operators in the team must be done (Figure 2). These individual estimates are then used to consider adaptations to suite the cognitive needs of each of the team members.

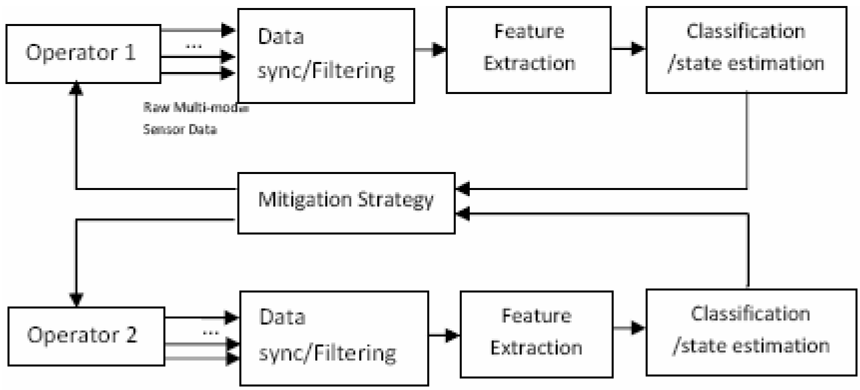


Fig. 2. Test-bed based on information processing model

The differences between these two models as seen from figures 1 and 2 are, in social cybernetic model, a single estimation of compliance metric (SPC) drives the mitigation, and same mitigation is applied to all the members of the team. In the information processing model, cognitive state estimations from all the operators is evaluated separately by the mitigation engine and adaptations are made reflecting the current need of each operator. As suggested by Henning et al. [11], using the social cybernetic model, a display of trajectory of SPC over time in a shared display to all the operators could also render helpful in achieving high over all team performance.

The model based on social cybernetics (Figure 1) would be only effective when all the operators are performing the same kind of synchronized task (such as, the laboratory experiment in section 2, in which two team members guided a virtual heavy object through a maze). In situations which require the operators to perform mutually exclusive tasks, the SPC might not be a good indicator of the overall team performance. In such cases the two models described above could be augmented where the SPC would become a part of individual cognitive estimates, and it could be used by the mitigation engine at appropriate times.

6 Conclusion and Future Work

In this paper we have sketched a test-bed framework for testing two models of cognition (social cybernetics and information processing). Each of these models holds contrasting views about cognition. Testing the usability of these models in terms of task specificity is essential before deploying any model in operational environment. In our future work, we intend to design experiments that involve tasks that require team members to perform compensatory actions (i.e. similar to maneuvering a virtual object in a maze) and compare them to tasks requiring members to perform mutually exclusive actions. Our goal then is to analyze the physiological data and find if there exists any relationship between the type of cognition model and team performance.

References

1. Mathieu, J., Heffner, T., Goodwin, G., Salas, E., Cannon-Bowers, J.: The influence of shared mental models on team process and performance. Journal of Applied Psychology 85(2), 273–283 (2000)
2. Kennedy, W., Trafton, J.: Using Simulations to Model Shared Mental Models. In: Proceedings of the Eighth International Conference on Cognitive Modeling. Taylor & Francis Psychology Press, Ann Arbor (2007)
3. Lee, M., Johnson, T., Lee, Y., O'Connor, D., Mohammed, K.: The Conceptual Framework of Factors Affecting Shared Mental Model. In: 27th Association for Educational Communications and Technology, Chicago, IL (2004)
4. Fiore, S., Salas, E.: Why we need team cognition, in Team Cognition: Understanding the Factors that Drive Process and Performance. In: Fiore, S., Salas, S. (eds.) American Psychological Association, Washington, DC, pp. 235–248 (2004)
5. Mohommed, S., Klimoski, R., Rentsch, J.: The Measurement of Team Mental Models: We Have No Shared Schema. Organizational Research Method 3(2), 123–165 (2000)
6. Kiekel, P., Cooke, N., Foltz, P., Shope, S.: Automating Measurement of Team Cognition through Analysis of Communication Data. In: Smith, M., et al. (eds.) Usability Evaluation and Interface Design, pp. 1382–1386. Lawrence Erlbaum Associates, Mahwah (2001)
7. Henning, R., Boucsein, W., Gil, M.: Social-physiological compliance as a determinant of team performance. International Journal of Psychophysiology 40, 221–232 (2001)
8. Smith, T., Henning, R.: Cybernetics of Augmented Cognition as an alternative to Information Processing. In: Schmorrow, D. (ed.) Foundations of Augmented Cognition, pp. 641–650. Lawrence Erlbaum Associates, Mahwah (2005)
9. Henning, R., Boucsein, W., Fekieta, R., Gil, M., Li, Q., Pratt, J.: Team biocybernetics based on social psychophysiology. In: Proceedings of PIE 2000, San Diego, CA (2000)
10. Vartak, A., Fidopiastis, C., Nicholson, D., Mikhael, W., Schmorrow, D.: Cognitive State Estimation for Adaptive Learning Systems using Wearable Physiological Sensors. In: First International Conference on Biomedical Electronics and Devices (Biosignals 2008), Funchal, Madeira, Portugal, pp. 147–152 (2008)
11. Henning, R., Smith, T., Korbelak, K.: Social psyphyiological compliance as a gauge of teamwork. In: Schmorrow, D. (ed.) Augmented Cognition, pp. 1228–1238. Lawrence Erlbaum Associates, Mahwah (2005)
12. Raley, C., Stripling, R., Kruse, A., Schmorrow, D., Patrey, J.: Augmented Cognition Overview: Improving Information Intake Under Stress. In: Proceedings of the Human Factors and Ergonomics Society (2004)
13. Wright, M., Kaber, D.: Effects of Automation of Information-Processing Functions on Teamwork. Human Factors 47(1), 50–66 (2005)
14. Hart, S., Staveland, L.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, pp. 139–183. Elsevier Science North Holland, Amsterdam (1988)

Measuring Cognitive Workload in Non-military Scenarios Criteria for Sensor Technologies

Jörg Voskamp and Bodo Urban

Fraunhofer-Institut für Graphische Datenverarbeitung, Institutsteil Rostock, Germany
{joerg.voskamp,bodo.urban}@igd-r.raunhofer.de
http://www.igd-r.fraunhofer.de/uat

Abstract. Augmented Cognition manifesting in the DARPA project is becoming of more and more interest to non-military application areas. First areas it is going to be applied are in flight control and power plant control. Measuring cognitive workload in the context of Augmented Cognition is bound to the application of sensor technologies and frameworks which are going to be applied to users. It is necessary to make Augmented Cognition Application in non-military areas as comfortable to the user as possible as we do not want to disturb her but to support her in her tasks. In this paper we will define criteria to be considered when designing Augmented Cognition applications in non-military environments.

Keywords: Augmented Cognition, Application, Sensors systems, sensor chriteria.

1 Introduction and Objective

Augmented Cognition is becoming more and more of interest to non-military application areas. First areas it is going to be applied are flight control centers, power plant control centers or simulation centers. Extending this thought one might think of other control scenarios where users and computer have to work collaboratively together, where tasks are to be carried both by men and machine, one supporing the other.

As Schmorrow and Kruse discussed "Augmented Cognition distinguished from its predecessors by the focus on the real-time feedback cognitive state of the user, as assessed through modern neuro-scientific tools." [1] Systems developed by using Task-Centered Design need to monitor both information flow as well as the cognitive state of the user to meet the objectives they are designed for [2].

This shortly leads to the application of Augmented Cognition in everyday life scenarios like buying train tickets at ticket machines, ATM usage, or writing articles, email and reports, programming new software or doing technology supported workplace learning [3]. As people in todays business life do not work on one task only at a time but many tasks simultaneously support is needed [4].

We are aiming at introducing Augmented Cognition technologies into different application areas and are looking for sensor types which can be used to recognize

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 304–310, 2009.

the users cognitive workload in a way the user would accept and provide the computer with the information it needs to support the user as best as possible. So the objective of the research presented here is to find criteria for sensor types in the respective environments and tasks acceptable by the users in their distinguished situations.

2 Measuares for of Cognitive Workload

As in military applications both "high demand" and "vigilance" scenarios are to be supported in adaptive automation application [5]. Application shall support users in their task work by optimizing her workload for optimal performance [6,7,8].

As measuring cognitive workload unfortunately is not possible directly, we have to use tools to recognize the humans load. There are different methods For measuring cognitive workload mostly applicable in laboratory environments. First subjective workload measures like Cooper-Harper-Scale [9], NASA task load index [10] or Subjective Workload Assessment Technique [11] and performance-based measures [12,13] have been proposed. Later estimation methods using psychophysiological parameter and body observations (e.g. Index of cognitive activity [14]) which would support the criteria on in-time estimation were developed. We focus on the estimation of cognitive workload by analyzing psychophysiological parameters and other data acquirable by sensors from the user.

3 Criteria for Sensor Technologies

Depending on the application area different criteria have to be met by the sensors. As they are the basis for Augmented Cognition applications they need to be accepted by the operators themselves [15,12]. Therefor the sensors need to be:

non-intrusive. Users will not carry implants or needles or other devices which may hurt them in any way [16].
non-obtrusive. The applied sensors do not disturb the handling of the user during the tasks performance.
easily applicable. The sensors are easily applicable to the user or work in remote sensing way, so would n't take much effort to start gathering data.

In technological terms sensors need to be usable for calculating estimates of the cognitive workload of the user. Technologically sensors should provide adequate data precision, should have an adequate data rate, and should be easily combinable with other sensors. Technological requirements are:

adequate precision. Sensors should be able to deliver data in a quality high enough to calculate estimates of the users cognitive workload.
adequate data rate. Data rate and transmission rate from sensors should be adequate to calculate estimates of cognitive workload.

combinability. Sensors support industrial and other standards, which allow them in any combination and connect them easily to a data transmission and collection framework [17,18].

Those criteria apply to all technical components, meaning sensors as well as data transmission components and computers necessary for cognitive workload estimation. When designing user friendly systems and applications those criteria should be taken into consideration. It is recommended to apply methods of human-centered and task-centered design [2].

Furthermore the factors sensitivity, diagnosticity, primary task intrusion, implementation requirements and selectivity should be taken care of [15,12]. An overview of sensor types vs. sensitity, obtrusiveness, avalibilty is given in [19].

4 Sensor Types and Classification

Science reports on research attempting to connect workload indices to one psychophysiological parameter. Unfortunately the

> "...indirect nature of derived psychophysiological parameters prevents a straightforward interpretation concerning the functional aspects of the human organic system..." [20]

– as it prevents from directly concluding from measured signals to cognitive workload. The main reason is that there is not a direct physical or physiological connection as the measured psychophysiological parameters finally just are reactions of the human brain and body on external and internal exposure and their resulting workload. Therefore it is absolutely necessary to receive high quality data of the measured psychophysiological signals. Advises on ensuring high quality signal data can be found in the references (e.g. [20]).

Estimates of the current cognitive workload can be calculated using combinations of measurements of different sensor types. When following up those criteria we may classify different sensor types in respect to their usability in new application areas and their acceptance by users. Different sensor types may be used and combined to acquire information on the users cognitive workload [21,12]. This approach works well in the field of emotion recognition [22] and accounts for measuring cognitive workload, as well. The following sensors have been proved to support reognition of cognitive workload:

- Electroencephalography (EEG), Magnetoencephalography (MEG) [23,24]
- Functional Near Infrared Neuroimaging (FNIR) [25,26]
- Pupil diameter using eye tracker [27,28,29,2,14,30,31,32]
- Psycho-physiological sensors like galvanic skin respondance, heart rate, blood pressure [23,33,34,35]
- Eye blink [30,28,36,34,8,37,27,32]
- Facial skin temperature [38,30]

– Other sensors types including selfmonitoring methods [39]
– Ad hoc wireless Body area network [17]

Sensors and technologies used to estimated users emotions might be applicable as well like for example facial play or voice and speech anaylysis [40,41].

Table 1 shows the sensor types and their respective criteria quality.

Table 1. Overview over sensortypes in respect to criteria

Sensor name	*sensor type*	*non-intrusive*	*non-obtrusive*	*easily applicable*	*adequate precision*	*adequate data rate*	*combinability*
EEG / MEG	directly connected to body	−−	−−	−	○	+	+
Near infrared	directly connected to body	−−	−−	−	○	+	+
Pupil diameter	remote sensor	++	+	○	−	−	+
Psycho-physiology	directly connected to body	−−	−−	−	○	+	+
Eye blink	remote sensor	++	+	○	−	−	+
Facial skin temperature	remote sensor	++	+	○	−	−	○

Looking at those criteria one finds out that remotely working sensors which automatically track the user during her task performance are in favor to the user acceptance criteria, whereas the sensors directly connected to the users body deliver the necessary data quality and data rate. All sensor types can be combined when using data wrapping technologies.

5 Conclusion, Outlook

We focus at applying Augmented Cognition in new application areas like Interactive electronic technical documentation[1], adaptive electronic tutorials and learning materials as well as adaptive user interfaces. User-centered technologies appraising cognitive workload will be accepted by the user if the sensing part is acceptable. As we showed those sensors have to be remotely operational, non-intrusive and non-obtrusive. Therefor more effort has to be spend on new and further development of tracking technologies remotely working. One new technology could for example include the users activity in the estimation of cognitive workload [42].

One key criterium to be considered furthermore in prospective investigations is the users privacy. Being able to track a users cognitive workload remotely may conflict with her idea of privacy.

[1] IETM – Interactive Electronic Technical Manual, for explanation see `http://en.wikipedia.org/wiki/IETM` or `http://www.cpt.fsu.edu/pdf/ietm1.pdf`

Even though the research area Augmented Cognition is still looking for its "killer app" [4], it is leaving the military sector and extends to application areas in every day life. In this paper we discussed the use and the combination of different sensor types for capturing user data relevant to her cognitive workload. We proposed several criteria which might be useful to select sensor types for new applications of Augmented Cognition.

References

1. Kruse, A.A., Schmorrow, D.D.: Session overview: Foundations of augmented cognition. In: Salvendy, G. (ed.) HCI International 2005 - 11th International Conference on Human-Computer Interaction, Caesars Palace, Las Vegas, Nevada, USA, July 22 - 27, vol. 11. Erlbaum, Mahwah (2005)
2. Orden, F.K., Viirre, E., Kobus, A.D.: Augmenting task-centered design with operator state assessment technologies. In: [43], pp. 212–219
3. Oertel, K., Kaiser, R., Voskamp, J., Urban, B.: Affectix - an affective component as part of an e-learning-system. In: [43], pp. 385–393
4. Forsythe, C., Berka, C., Matthews, R., Wagner, J.: Making the giant leap with augmented cognition technologies: What will be the first 'killer app'? In: [43], pp. 434–438
5. Prinzel, J.L., Freeman, G.F., Scerbo, W.M., Mikulka, J.P., Pope, T.A.: A closed-loop system for examining psychophysiological measures for adaptive task allocation. International Journal of Aviation Psychology 10(4), 393–410 (2000)
6. Scerbo, W.M., Freeman, G.F., Mikulka, J.P.: A biocybernetic system for adaptive automation. In: [44], pp. 241–253
7. Ward, R.D., Marsden, P.H.: Physiological responses to different web page designs. Int. J. Hum.-Comput. Stud. 59(1-2), 199–212 (2003)
8. Veltman, A.J., Jansen, C.: The role of operator state assessment in adaptive automation (2006)
9. Cooper, E.G., Harper, P.R.: The use of pilot ratings in the evaluation of aircraft handlicng characteristics, Washington, D.C . (1969)
10. Hart, G.S., Staveland, E.L.: Development of nasa-tlx(task load index) resultsof empirical and theoretical research. In: [45]
11. Reid, B.G., Nygren, T.E.: The subjective workload assessment technique: a scaling procedure for measuring mental workload. In: [45], pp. 185–215
12. Embrey, D., Blackett, C., Marsden, P., Peachey, J.: Development of a human cognitive workload assessment tool (2006)
13. Freude, G., Ullsperger, P.: Slow brain potentials as a measure of effort? applications in mental workload studies in laboratory settings. In: [44], pp. 255–267
14. Marshall, P.S.: The index of cognitive activity: Measuring cognitive workload. In: Persensky, J.J. (ed.) New century, new trends - Proceedings of the 2002 IEEE 7th Conference on Human Factors and Power Plants, Scottsdale, Arizona, New York, NY, September 15 - 19, pp. 7.5–7.9. Institute of Electrical and Electronics Engineers (2002)
15. O'Donnel, D.R., Eggemeier, T.F.: Workload assessment methodology. In: Boff, R.K. (ed.) Handbook of perception and human performance, vol. 2, pp. 42/1–42/49. A Wiley-Interscience publication, New York (1986)
16. Guhe, M., Liao, W., Zhu, Z., Ji, Q., Gray, D.W., Schoelles, J.M.: Non-intrusive measurement of workload in real-time. Human Factors and Ergonomics Society Annual Meeting Proceedings 5, 1157–1161 (2005)

17. Ikehara, S.C., Biagioni, E., Crosby, E.M.: Ad-hoc wireless body area network for augmented cognition sensors. In: [43], pp. 38–46
18. Mader, S., Peter, C., Göcke, R., Schultz, R., Voskamp, J., Urban, B.: A freely configurable, multi-modal sensor system for affective computing. In: André, E., Dybkjær, L., Minker, W., Heisterkamp, P. (eds.) ADS 2004. LNCS, vol. 3068, pp. 313–318. Springer, Heidelberg (2004)
19. Grootjen, M., Neerincx, A.M., Weert, C.J.: Task-based interpretation of operator state information for adaptive support. In: Schmorrow, D.D., Stanney, M.K., Reeves, M.L. (eds.) Foundations of Augmented Cognition, 2nd edn., Strategic Analysis, Arlington, Virginia, pp. 236–242 (2006)
20. Luczak, H., Göbel, M.: Signal processing and analysis in application. In: [44], pp. 79–110
21. Greef, T., Dongen, K., Grootjen, M., Lindenberg, J.: Augmented cognition: Reviewing the symbiotic relation between man and machine. In: [43], pp. 439–448
22. Castellano, G., Kessous, L., Caridakis, G.: Emotion recognition through multiple modalities: Face, body gesture, speech. In: Peter, C., Beale, R. (eds.) Affect and Emotion in Human-Computer Interaction, pp. 92–103. Springer, Berlin (2008)
23. Ribback, S.: Psychophysiologische Untersuchung mentaler Beanspruchung in simulierten Mensch-Maschine-Interaktionen. PhD thesis, Universität Potsdam, Potsdam (2003)
24. Belyavin, A., Ryder, C., Dickson, B.: Development of gauges for the qinetiq cognition monitor. In: [43], pp. 3–12
25. Izzetoglu, M., Izzetoglu, K., Bunce, S., Ayaz, H., Devaraj, A., Onaral, B., Pourrezaei, K.: Functional near-infrared neuroimaging. IEEE Transactions on Neural Systems and Rehabilitation Engineering 13(2), 153–159 (2005)
26. He, P., Yang, B., Hubbard, S., Estepp, J., Wilson, G.: A sensor positioning system for functional near-infrared neuroimaging. In: [43], pp. 30–37
27. Klingner, J., Kumar, R., Hanrahan, P.: Measuring the task-evoked pupillary response with a remote eye tracker. In: Räihä, K.J., Duchowski, T.A. (eds.) Proceedings of the 2008 symposium on Eye tracking research & applications, pp. 69–72. ACM, New York (2008)
28. Tsai, Y.F., Viirre, E., Strychacz, C., Chase, B., Jung, T.P.: Task performance and eye activity - predicting behavior relating to cognitive workload. Aviation, Space, and Environmental Medicine 78(suppl. 1), B176–B185 (2007)
29. Ahlstrom, U., Frieman-Berg, J.F.: Using eye movement activity as a correlate of cognitive workload. International Journal of Aviation Psychology 36(7), 623–636 (2006)
30. Wang, L.M., Duffy, G.V., Du, Y.: A composite measure for the evaluation of mental workload. In: Duffy, V.G. (ed.) HCII 2007 and DHM 2007. LNCS, vol. 4561, pp. 460–466. Springer, Heidelberg (2007)
31. Zhang, Y., Owechko, Y., Zhang, J.: Driver cognitive workload estimation - a data-driven perspective. In: The 7th International IEEE Conference on Intelligent Transportation Systems Proceedings, Piscataway, NJ, pp. 642–647. IEEE Operations Center (2004)
32. Sirevaag, J.E., Stern, A.J.: Ocular measures of fatigue and cognitive factors. In: [44], pp. 269–287
33. Byrne, A.E., Parasuraman, R.: Psychophysiology and adaptive automation. Biological Psychology 42(3), 249–268 (1996)
34. Veltman, A.J., Gaillard, W.A.: Physiological workload reactions to increasing levels of task difficulty (1996-06-21)

35. Lenneman, K.J., Backs, W.R.: The validity of factor analytically derivced cardiac autonomic components for mental workload assessment. In: [44], pp. 161–175
36. Boehm-Davis, A.D., Gray, D.W., Schoelles, J.M.: The eye blink as a physiological indicator of cognitive workload. Human Factors and Ergonomics Society Annual Meeting Proceedings 5, 116–119 (2006)
37. Heishman, R., Duric, Z.: Using eye blinks as a tool for augmented cognition. In: [43], pp. 84–93
38. Or, K.L.C., Duff, G.V.: Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement. Occupational Ergonomics 7(2), 83–94 (2007)
39. Picard, W.R., Liu, K.K.: Relative subjective count and assessment of interruptive technologies applied to mobile monitoring of stress. International Journal Of Human-Computer Studies 65(4), 361–375 (2007)
40. Göcke, R., Voskamp, J.: Optische mimikanalyse. In: Hambach, S., Urban, B. (eds.) Multimedia & Bildung - Beiträge zu den 4. IuK-Tagen Mecklenburg-Vorpommern, pp. 192–196. Fraunhofer-IRB-Verl., Stuttgart (2003)
41. Truong, P.K., Leeuwen, A.D., Neerincx, A.M.: Unobtrusive multimodal emotion detection in adaptive interfaces - speech and facial expressions. In: [43], pp. 354–363
42. Bieber, G., Peter, C.: Using physical activity for user behavior analysis. In: 1st ACM International Conference on PErvasive Technologies Related to Assistive Environments (2008)
43. Schmorrow, D.D., Reeves, M.L. (eds.): HCII 2007 and FAC 2007. LNCS (LNAI), vol. 4565. Springer, Heidelberg (2007)
44. Backs, W.R., Boucsein, W. (eds.): Engineering Psychophysiology - Issues and applications. Erlbaum, Mahwah (2000)
45. Hancock, A.P., Meshkati, N. (eds.): Human Mental Workload. Advances in psychology, vol. 52. North-Holland, Amsterdam (1988)

Combined Effects of Sleep Deprivation, Narrow Space, Social Isolation and High Cognitive Workload on Cognitive Ability of Chinese Operators

Yijing Zhang[1,2], Xueyong Liu[2], Zhizhong Li[1], Bin Wu[2,*], Fang Liu[2], Xiaolu Jing[2], Jun Wang[2], Haibo Qin[2], and Su Wu[1]

[1] Department of Industrial Engineering, Tsinghua University, Beijing, 100084, P.R. China
[2] China Astronaut Research and Training Center, Beijing, 100094, P.R. China
wubinacc@sina.com

Abstract. This study aims to investigate the effects of sleep deprivation on the cognitive abilities of Chinese subjects under a combined scenario of narrow space, social isolation, and high cognitive workload. Twelve subjects participated in the experiment of 72-hour sleep deprivation, and took 15 cognitive ability tests for three times at the first day of sleep deprivation (level 1), the second day of sleep deprivation (level 2) and the third day of sleep deprivation (level 3) respectively. The result data analyses show that: most of the cognitive abilities do not change significantly, but the value of special graphics search is increased significantly with the increasing of sleep deprivation time (p=0.01). In addition, when four cognitive ability tests were combined into one complex measure, the effect of sleep deprivation becomes more significant. The results mainly support the Hockey's compensatory control model. And they may be due to two other reasons: one is that the combined stressors will counteract each other; the other is that learning effect can improve operators' performance. The results also imply that sensitive and complex measures need to be developed and used to reflect the compound effects of sleep deprivation under such combined situation.

Keywords: Sleep deprivation; narrow space; social isolation; high cognitive workload; cognitive ability.

1 Introduction

Realized in a large number of anecdotal reports and observations, astronauts' cognitive performance in spaceflight may be seriously influenced due to the extreme working and living conditions in space, such as microgravity, narrow space, social isolation, sleep disturbance, and high cognitive loading [1, 2]. Two factors which may impair the cognitive and psychomotor performance of astronauts were distinguished by Hockey [1, 3]: the direct effects of microgravity on specific brain mechanism and non-specific stressors such as cumulative sleep loss, workload, or the physical and emotional burden of adapting to the conditions in space. Among those non-specific

* Corresponding author.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 311–316, 2009.

stressors, sleep disturbances in spaceflight was recognized as one of the most important factors contributing to impaired performance of astronauts [4].

The possible effects of sleep restriction or sleep deprivation on cognitive performance were studied [5-8]. Some studies suggest that restrictions of sleep may result in cognitive performance decrements, which include increased response times and number of lapses in simple reaction time tasks, slowing of performance in mental arithmetic tasks, or impaired working memory functions. Russo also indicated that acute sleep deprivation degraded visual perceptual and simple motor performance [8]. However, some experiment results also supported the compensatory theory [9-11]. It was also found that the effects of sleep deprivation might be attenuated as the cognitive demands of a task increased [3, 12]. Sleep deprivation may cause some performance impairment which does not show apparently [13].

From the above studies, it can be noticed that sleep deprivation has effects on cognitive ability with complicated phenomena. Through one 48 hours sleep deprivation experiment, May et al concluded that the cognitive abilities requiring high response speed and high attention were obviously damaged by sleep deprivation, and those which included much intellectual element were not depressed but enhancive during the sleep deprivation [14]. It was also found that the effects of sleep deprivation might be attenuated as the cognitive demands of a task increased [15].

Although many studies have been conducted on the non-specific stressors, most of them considered sleep deprivation as a single factor, without paying attention to the combined effects of the non-specific stresses. Additionally, there is very limited knowledge about the cognitive ability of Chinese population under such extreme conditions. In this study, an experiment was designed to investigate the effects of sleep deprivation on the cognitive abilities of Chinese subjects under a combined scenario of narrow space, social isolation and high cognitive workload.

2 The Experiment

This experiment was a single-factor experiment, in which the subjects were asked to keep awake for 72 hours in an isolated room.

2.1 Subjects

Twelve male subjects were selected from twenty students of China Agricultural University. Their ages ranged from 20 to 26 years (M =22.3, SD =1.82). All subjects passed a general physical examination and a mental health evaluation by the Eysenck Personality Questionnaire (EPQ) and the Symptom Checklist 90 (SCL-90). The twelve subjects were divided into four groups randomly and equally. One group with three subjects attended the experiment at a time. This is the standard size for a Chinese spaceship crew. However, unlike the normal space flight, the subjects in a group had very limited social contact with each other, but focused on their experiment tasks.

2.2 Independent Variables

Sleep deprivation was considered as an independent variable in this experiment, which lasted for approximately 72 hours (over a four-day span). Three levels of sleep deprivation were examined: level 1 (the first day), level 2 (the second day), and level 3 (the third day). During the experiments, subjects were monitored by video cameras and waken up by a ring if they tried to sleep.

2.3 Fixed Variables

The three fixed variables, narrow space, social isolation and high cognitive workload, were tested in this experiment.

The experiment room was small and isolated, the wall of which was sound and magnetic signal proofing. During the experiment, all the direct channels for information exchange were cut off. Necessaries were delivered through a double-layer window. All subjects performed experiment tasks according to the experiment schedule. In order to monitor the experiment process and keep the subjects awake, video cameras and loudspeakers were equipped in the room.

For giving high cognitive workload, various cognitive tasks were assigned to the subjects throughout the experiment. Beside those cognition tasks, spaceflight task operations, physiological measurements, emotion evaluations, and personality measurements were also performed to check the status of the subjects. There were multiple research objectives in this experiment. In this paper, we focused on the 15 cognitive ability tests and the analyses on the other data were not reported.

2.4 Dependent Variables

During the experiment, 15 cognitive ability tests were measured three times at the first day of sleep deprivation (level 1), the second day of sleep deprivation (level 2) and the third day of sleep deprivation (level 3), respectively. The 15 tested cognitive ability items were memory search, rule finding, direction discrimination, symbol substitution, spatial rotation, speed estimation, combined graph, comparative scale, digit search, mental rotation, symbol discrimination, attention span, special graph search, analogic test, and hiding graph. The value of each cognitive ability test ranges from 1 to 10.

The 15 cognitive ability items were tested with a computer software toolkit for psychology measurement, which has well designed and tested in astronaut selection and training.

2.5 Experiment Procedure

The experiment was conducted in China Astronaut Training Centre and had three phases. The first phase was subject screening. Through both medical examination and mental health screening, 12 subjects were selected from 20 volunteers and were admitted to participate in the formal experiment after they signed informed consent forms.

The second phase was experiment preparation which began for one to three days before the experiment was started. First, the subjects provided their personal information such as gender, age, and education background. Then they listened to instructions of the experiment, and then practised all the 15 cognition tests and other required operations.

The third phase was the formal sleep deprivation experiment which was started at 9:00 am of the first day and ended at the same time of the forth day. The subjects took the 15 cognitive ability tests once on each day.

3 Results

SPSS 15.0 was used for the data processing in this study. The effects of sleep deprivation at the 15 cognitive ability tests were analyzed. Four of the 15 items passed both the normality and homogeneity of variances tests and thus One-way ANOVA method was applied, including memory search, rule finding, direction discrimination, and symbol substitution. The other items failed to pass the normality and homogeneity of variances tests. Hence, nonparametric tests (Friedman test) were applied for those data. Table 1 shows the results.

From Table 1 it can be seen that most of the cognitive ability do not significantly change during the four days. However, the value of special graphics search is significantly increased as the sleep deprivation time passed (p=0.01). Among these cognitive tests, some show a decreasing trend while others show an increasing trend.

Though the 15 tests were designed to test different cognitive abilities, some of them could measure the same cognitive function. For example, memory search, symbol substitution, digit search and symbol discrimination all reflect short-term memory. To study the effect of sleep deprivation on short-term memory, the sum of the four test values were analyzed. The sum did not pass the normality and homogeneity of variances tests, and thus Friedman test was applied. It show that the short-term memory (combined tests) has a marginally significant decrease (p = 0.098).

4 Discussion

As a whole, these results support the Hockey's compensatory control model which means operators have the mechanism of maintaining performance [9, 10]. The results may be due to another two reasons in this experiment. The first is the interaction of various factors. Those factors that include narrow environment, isolation, sleep deprivation and high cognitive workload had shown significant effects on cognitive ability impairment separately; however, when they appear together, the results may turn different, e.g. the tiredness caused by sleep deprivation might be reduced by the high cognitive workload. The second reason is the learning effect. Although the subjects were well trained before the formal experiment, the learning effect can not be totally avoided.

In addition, it should be noticed that sleep deprivation has different effects on different cognitive abilities. Those cognitive abilities with low cognitive load show a decreasing trend, such as digit search; while those requiring high cognitive demand show an increasing trend, such as special graph search. In this experiment, each of the four tests reflecting short-term memory individually show a decreasing trend, but it is not statistically significant. When they were combined into one complex measure, the effect of sleep deprivation becomes more significant. It implies that more sensitive and complex cognition measurements need to be developed to reflect the compound effects of sleep deprivation under such combined situation.

Table 1. The analysis result of 15 items of cognitive ability

Cognition measures	1st day		2nd day		3rd day			
	M	SD	M	SD	M	SD	F(2, 33)	*p*
Memory search	6.48	1.28	6.45	1.42	6.15	0.903	0.132	0.877
Rule finding	6.04	1.30	6.13	1.35	6.30	1.48	0.094	0.910
Direction discrimination	6.06	3.07	6.43	2.74	6.39	2.40	0.067	0.935
Symbol substitution	6.72	1.41	6.25	1.56	5.86	1.77	0.879	0.425
Spatial rotation [a]	0.49	0.056	0.47	0.072	0.52	0.081	0.110	0.897
Speed estimation [a]	0.61	0.065	0.62	0.046	0.51	0.067	0.881	0.424
Combined graph [b]	7.08	0.446	7.55	0.48	7.81	0.420	3.95	0.139
Comparative scale [b]	7.86	0.420	7.93	0.552	7.74	0.594	1.62	0.446
Digit search [b]	7.98	0.509	7.68	0.58	6.78	0.971	3.55	0.17
Mental rotation [b]	5.48	0.774	6.35	0.71	4.93	0.872	2.84	0.242
Symbol discrimination [b]	8.63	0.269	8.34	0.224	8.09	0.432	2.48	0.289
Attention span [b]	8.57	0.339	8.59	0.285	8.34	0.286	0.667	0.717
Special graph search [b]	5.35	0.811	5.14	0.721	7.13	0.634	9.19	0.01
Analogic test [b]	8.49	0.291	8.38	0.309	8.23	0.324	1.08	0.582
Hiding graph [b]	8.30	0.294	8.27	0.321	8.24	0.353	1.70	0.428
Short-term memory (combined tests) [b]	29.8	4.07	28.7	4.09	26.9	5.70	4.64	0.098

Note: a - Logarithm conversion was applied on the original data. b - Friedman Test.

5 Conclusions

This study tries to explore the effects of sleep deprivation on the cognitive ability of the Chinese under a combined situation. Though it is not found that sleep deprivation combined with other performance influencing factors would lead to significant cognitive impairments for Chinese operators, the effect of sleep deprivation exist in reality and cannot be neglected.

Acknowledgments. This work was supported in part by Military Foundation for Medicine and Health Research (06MA002), and the Manned Space Project of China.

References

1. Kanas, N., Manzey, D.: Space Psychology and Psychiatry, 1-87. Kluwer Academic Publishers, Dordrecht (2004)
2. Kubis, J.F., McLaughlin, E.J., Jackson, J.M., Rusnak, R., McBride, G.H., Saxon, S.V.: Task and work performance on Skylab missions 2, 3, and 4: time and motion study - experiment M151. In: Johnston, R.S., Dietlein, L.F. (eds.) Biomedical results from Skylab (NASA-SP377), pp. 136–154. National Aeronautics and Space Administration, Washington (1977)
3. Hockey, G.R.J.: Changes in operator efficiency as a function of environmental stress, fatigue, and circadian rhythms. In: Boff, K.R., Kaufman, L., Thomas, J.P. (eds.) Handbook of Perception and Performance, Vol II: Cognitive Processes and Performance, pp. 41.1–41.49 . Wiley, New York (1986)
4. Nechaev, A.P.: Work and rest planning as a way of crew member error management. Acta Astronautica 49, 271–278 (2001)
5. Belenky, G., Wesensten, N.J., Thorne, D.R., Thomas, M.L., Sing, H.C., Redmond, D.P., Russo, M.B., Balkin, T.J.: Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. Journal of Sleep Research 12, 1–12 (2003)
6. Van-Dongen, H.P., Maislin, G., Mullington, J.M., Dinges, D.F.: The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. Sleep 26(2), 117–126 (2003)
7. Dinges, D.F., Pack, F., Williams, K., Gillen, K.A., Powell, J.W., Ott, G.E., Aptowicz, C., Pack, A.I.: Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. Sleep 20(4), 267–277 (1997)
8. Russo, M., Sing, H., Kendall, A., Johnson, D., Santiago, S., Escolas, S., Holland, D., Thorne, D., Hall, S., Redmond, D., Thomas, M.: Visual Perception, Flight Performance, and Reaction Time impairments in Military Pilots during 26 Hours of Continuous Wake: Implications for Automated Workload Control Systems as Fatigue Management Tools. In: Strategies to Maintain Combat Readiness during Extended Deployments – A Human Systems Approach. Meeting Proceedings RTO-MP-HFM-124, pp. 1–16. RTO(Paper 27). Neuilly-sur-Seine, France (2005)
9. Hockey, G.R.J., Wastell, D.G., Sauer, J.: Effects of sleep deprivation and user interface on complex performance: a multilevel analysis of compensatory control. Human Factors 40(2), 233–253 (1998)
10. Hockey, G.R.J.: Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework. Biological Psychology 45, 73–93 (1997)
11. Lichacz, F.M.J.: Examining the effects of combined stressors on dynamic task performance. International Journal of Aviation Psychology 15(1), 45–66 (2005)
12. Harrison, Y., Horne, J.A.: The impact of sleep deprivation on decision making: a review. Journal of Experimental Psychology: Applied 6(3), 236–249 (2000)
13. Grugle, N.L.: Understanding the effects of sleep deprivation on executive function, complex task performance and situation awareness. PHD Dissertation, Virginia Polytechnic Institute and State University 94, 141 (2005)
14. May, J., Kline, P.: Measuring the effects upon cognitive abilities of sleep loss during continuous operations. Br. J. Psychol. 78(4), 443–455 (1987)
15. Harrison, Y., Horne, J.A.: The impact of sleep deprivation on decision making: a review. Journal of Experimental Psychology: Applied 6(3), 236–249 (2000)

Quantifying the Feasibility of Compressive Sensing in Portable Electroencephalography Systems

Amir M. Abdulghani, Alexander J. Casson, and Esther Rodriguez-Villegas

Circuits and Systems Group, Department of Electrical and Electronic Engineering,
Imperial College London, UK, SW7 2AZ
{amirm,acasson,e.rodriguez}@imperial.ac.uk

Abstract. The EEG for use in augmented cognition produces large amounts of compressible data from multiple electrodes mounted on the scalp. This huge amount of data needs to be processed, stored and transmitted and consumes large amounts of power. In turn this leads to physically large EEG units with limited lifetimes which limit the ease of use, and robustness and reliability of the recording. This work investigates the suitability of compressive sensing, a recent development in compression theory, for providing online data reduction to decrease the amount of system power required. System modeling which incorporates a review of state-of-the-art EEG suitable integrated circuits shows that compressive sensing offers no benefits when using an EEG system with only a few channels. It can, however, lead to significant power savings in situations where more than approximately 20 channels are required. This result shows that the further investigation and optimization of compressive sensing algorithms for EEG data is justified.

Keywords: Compressive Sensing, Electroencephalogram, Power efficient, Wireless Systems.

1 Introduction

Augmented cognition systems which aim to close the loop on human-computer interactions intrinsically require some form of physiological monitoring of the human. The electroencephalogram (EEG), which places multiple recording electrodes on the head and records the micro-Volt sized signals produced, is a popular choice for this. The eventual level of end-user acceptance of augmented cognition technology will thus be strongly dependent on the miniaturization of the EEG technology so that it is discrete, comfortable and long-lasting. This last point is also an important factor in the design of robust systems. For example, in the dismounted solider scenario the EEG equipment may have to operate reliably over many days while the solider is out of contact with friendly forces. Also, when using EEG devices with people with learning difficulties, physically large systems requiring frequent battery changes could be a major impediment to producing reliable and repeatable results.

It has been shown [1] that power consumption, and in turn the battery size, is the major determining factor in the overall device size and system lifetime. For wireless

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 319–328, 2009.

EEG systems (which are potentially more discrete and wearable) most of the system power is consumed by the wireless transmitter, and thus it is desirable to compress the raw EEG data in real-time on the wearable device, in order to reduce the amount of data to transmit, and thus increase the operating lifetime or decrease the battery size.

This paper investigates applicability of compressive sensing, a recent development in compression theory, for this online data compression. An overview of compressive sensing theory is given in Section 2, but the work here assumes, based upon previous studies with EEG data [2] as well as in applications such as MRI where compressive sensing has been used very successfully [3], that compressive sensing can be used to achieve an acceptable compression ratio and reconstruction error. Instead, the focus here is on investigating the computational complexity of the method, and the implications of this for its implementation in an online, low power system.

Based upon the system modeling presented in Section 3, it is found that compressive sensing is not a beneficial compression technique when applied to an EEG system consisting of only a few channels, as commonly used in augmented cognition systems. However, as more channels are used, and many systems may commonly use 128 or more channels, the compressive sensing scheme can lead to a significant reduction in the overall power consumption. These results are presented, and the implications discussed, in Section 4.

2 Compressive Sensing Overview

The concept of compressive sensing [4] and [5] is based on the fact that there is a difference between the *rate of change* of a signal and the *rate of information* in the signal. Traditional Nyquist sampling, putting the signal into the digital domain ready for wireless transmission, is based on the former. The Nyquist theorem states that it is necessary to sample the signal at a rate at least twice the maximum rate of change present. A conventional compression algorithm would then be applied to all of these samples taken to remove any redundancy present, giving a reduced number of bits that represent the signal.

In contrast, compressive sensing exploits the information rate within a particular signal. Redundancy in the signal is removed during the sampling process itself, leading to a lower effective sampling rate. Provided certain conditions are satisfied [5], sampling at a sub-Nyquist rate the signal can still be accurately recovered.

To illustrate this, consider an EEG signal of interest $\mathbf{x}$ which is a vector of N digital samples; i.e. $x[n]$ where n =1, 2 … N. Then assume that this signal can be represented by a projection onto a different basis set:

$$x = \sum_{i=1}^{N} s_i \Psi_i \quad \text{or} \quad \mathbf{x} = \mathbf{\Psi s} \tag{1}$$

where $\mathbf{s}$ is a N×1 basis function vector and $\mathbf{\Psi}$ is a N×N basis matrix. The matrix $\mathbf{s}$ can be calculated from the inner product of $\mathbf{x}$ and $\mathbf{\Psi}$:

$$s_i = \langle \mathrm{x}, \Psi_i \rangle \,. \tag{2}$$

For example, if $\mathbf{\Psi}$ is the Fourier basis set of complex exponential functions, $\mathbf{s}$ is the Fourier transform of $\mathbf{x}$ and both $\mathbf{s}$ and $\mathbf{x}$ represent the signal equivalently, but in different domains. In compressive sensing $\mathbf{\Psi}$ is chosen so that $\mathbf{s}$ is sparse – a vector is K-sparse if has K non-zero entries and the remaining $N–K$ entries are all zero. $\mathbf{s}$ is thus a more *compact* representation of the signal than the original $\mathbf{x}$.

Similar to this projection, assume that $\mathbf{x}$ can be related to another signal $\mathbf{y}$:

$$\mathbf{y} = \mathbf{\Phi x} \tag{3}$$

where $\mathbf{y}$ is a $M\times1$ vector and $\mathbf{\Phi}$ is a matrix of dimensions $M\times N$ where $M<N$. Thus:

$$\mathbf{y} = \mathbf{\Phi \Psi s}\,. \tag{4}$$

Provided that $\mathbf{\Phi}$ is correctly chosen so that no significant information is lost during the reduction in dimensionality, it is possible to use $\mathbf{\Phi}$ to sample the sparse signal $\mathbf{s}$, rather than the original signal $\mathbf{x}$ to give an output vector $\mathbf{y}$ which has only M entries rather than the original N. If $M<N$ data compression is thus achieved, and the signal $\mathbf{y}$ would be transmitted from the portable EEG unit. It can be shown [5] that this technique is possible if $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are incoherent; that is if the elements of $\mathbf{\Phi}$ and $\mathbf{\Psi}$ have low correlation.

Given a compressed measurement $\mathbf{y}$ at the receiver, the signal $\mathbf{x}$ can be reconstructed by solving the L1 problem:

$$\min_{\mathbf{s}\in\Re^N} \|\mathbf{s}\|_{\ell 1} \text{ subject to } y_i = \langle \mathbf{\Phi}_\mathrm{i}, \mathbf{\Psi s}\rangle \tag{5}$$

which finds the vector $\mathbf{s}$ with the lowest L1 norm that satisfies the observations made. This is then easily converted back into $\mathbf{x}$. In general, the L1 minimization problem is non-trivial and computationally complex, but there is no need for this to run online in the portable EEG unit. The EEG signal $\mathbf{x}$ will be sampled as signal $\mathbf{y}$, and these samples wirelessly transmitted to a base station which will then regenerate $\mathbf{x}$ from $\mathbf{y}$ offline. The fact that compressive sensing based data compression has all of its computational complexity in the backend, where power and size constraints are not as stringent is a major factor motivating its investigation.

Previous work, [2], using Gaussian Random matrices with independent and identically distributed random variables or the Bernoulli matrix as the measurement matrix $\mathbf{\Phi}$ has shown promising (although not conclusive) results on the application of compressive sensing theory to EEG signals. However, the optimal choice of N and M, which set the amount of data compression but also reconstruction error, and the choice of optimization algorithm for the reconstruction are still open questions.

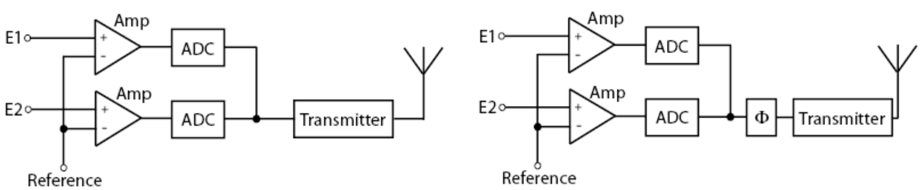
3 System Modeling and Feasibility Analysis Framework

The answering of these open questions is not the focus of this work. Before considering them it is instead essential to assess the feasibility of the overall scheme from the power point of view: the aim of compression must be to reduce the system power consumption so it is necessary to assess whether this is achievable. There is little practical point in optimizing the parameters identified above if this decrease is system power is not achievable.

An investigation of this can be carried out by considering the simplified EEG system model from Figure 1a. This incorporates an input instrumentation amplifier to amplify the small EEG signals from the head, an analogue-to-digital converter (ADC) to convert the EEG signals into the digital domain ready for transmission, and a transmitter. Given this, the system power consumption for a C channel system (P_{sys}) is given by:

$$P_{sys} = C\left(P_{Amp} + P_{ADC} + Jf_s R\right) \tag{6}$$

where P_x is the power consumption of block x from Figure 1a, f_s is the ADC sampling frequency, R the number of bits per sample and J the net transmission power per bit such that Jf_sR gives the transmitter power consumption. It is assumed that band-limiting of the EEG signal is incorporated into the instrumentation amplifier.



a) A standard EEG unit.

b) An EEG unit incorporating compressive sensing.

Fig. 1. Simplified EEG system model to enable power modeling

For comparison, Figure 1b illustrates the necessary modifications required to incorporate compressive sensing into the EEG system. The compressive sensing is implemented in the discrete domain and all that is required is a block to generate the measurement matrix $\mathbf{\Phi}$ which would be used to select a random set of samples to form $\mathbf{y}$. Elements in $\mathbf{\Phi}$ form a pseudo-random sequence following a particular probability distribution.

Given this, the system power consumption per channel (P_{sys_cs}) is now modified to:

$$P_{sys\_cs} = C\left(P_{Amp} + P_{ADC}\right) + P_{RNG} + P_{DSP} + P_{Sync} + J\left(\frac{M}{N} \cdot Cf_s R + S\right). \tag{7}$$

Here the instrumentation amplifier and ADC power consumptions are unchanged, but three extra terms representing the extra hardware required are also present: a random number generator (P_{RNG}) is used to generate the $\mathbf{\Phi}$ matrix; a DSP or microcontroller (P_{DSP}) is used to carry out the matrix multiplications from (4); and a synchronization unit (P_{Sync}) is used so that $\mathbf{\Phi}$ matrix does not need to be transmitted – it can be reconstructed at the receiver based upon the known pseudo-random sequence and a seed. Only one of each of these blocks is required regardless of the number of channels in the system. In addition to these blocks, the transmitter power consumption has changed in a number of ways.

Firstly, the power required to transmit the number of data bits (CJf_sR in (6)) has been reduced by a factor of M/N. This corresponds to the compressive sampling in (3) where there is a reduction in dimensionality between $\mathbf{x}$ and $\mathbf{y}$. In addition, however, it is necessary to also transmit S bits of extra data corresponding to the synchronization required between the EEG unit and the receiver to regenerate the $\mathbf{\Phi}$ matrix. Again the number of bits needed does not depend on the number of channels present as the same $\mathbf{\Phi}$ matrix will be used for all channels.

To assess the feasibility of compressive sensing based systems in low power portable EEG equipment it is thus simply a matter of comparing (6) and (7) using realistic, and state-of-the-art, figures. For this, five separate blocks need to be considered. These are discussed in turn below and the end figures used, incorporating some rounding and safety factors, are summarized in Table 3.

Instrumentation amplifier. The input amplifier is responsible for amplifying the small EEG signals detected on the scalp (typically in the range 2 µV to 500 µV) so that they match the input range of the analogue-to-digital converter. In addition it is assumed that the signal is band-limited (to the approximate range 0.5 Hz to 70 Hz) in this stage. The performance of a range of state-of-the-art integrated circuit EEG amplifiers is illustrated in Table 1.

Table 1. A comparison of state-of-the-art EEG suitable instrumentation amplifiers

Reference	[6]	[7]	[8]	[9]	[10]	[11]	[12]
Gain [dB]	40	44	77	40	44	48	38
Bandwidth [Hz]	30	1000	600	500	200	100	200
Input referred noise integrated over bandwidth [µVrms]	1.6	1.5	0.26	10	1.3	0.59	0.89
Process technology [µm]	1.5	0.35	1	3	0.35	0.5	3
Supply voltage [V]	2.5	1	5	2.5	1	3	2
Power consumption (P_{Amp}) [µW]	0.9	1.4	3,000	75	50	7	34

Analogue-to-digital converter. The ADC is responsible for digitizing the EEG ready for transmission, and the core parameter of interest is the resolution which sets the number of bits taken per sample and the end level of quantization noise. If any d.c. offset in the EEG signal is removed by the instrumentation amplifier a resolution of 10-12 bits is generally sufficient for the clinical recording of the EEG [13]. Given this, and the approximate 200 Hz sampling rate required, the performance of a selection of state-of-the-art ADCs is illustrated in Table 2. Thus, based upon Table 1 and Table 2, an overall power consumption for the instrumentation amplifier and the ADC of 2 µW is assumed to be reasonable.

Table 2. A comparison of state-of-the-art ADCs with suitable resolutions and sampling rates

Reference	[11]	[14]	[15]	[16]	[17]	[18]
Resolution (R) [bits]	11	8	12	10	10	10
Sampling rate (f_s) [kS/s]	8	1000	0.5	0.7	100	3.2
Process technology [μm]	0.5	0.18	0.18	0.8	0.09	0.5
Supply voltage [V]	3	0.6	1	2	0.65	1.2
Power consumption (P_{ADC}) [μW]	23	0.4	0.2	2.3	27	0.055

Random number generator. An example random number generator for use in generating the **Φ** matrix is given in [19], and as [19] also contains a comparison with other random number generators with respect to bit rates and power consumption, it is taken to be representative. This operates at 5V on a 0.35 μm process consuming 2.9 μW for an output data rate of 500 bps.

Processor unit for matrix multiplications. The matrix multiplications to carry out the compressive sensing will need to be implemented in either a dedicated digital signal processing chip or a microcontroller. The overall power of this depends strongly on the specifications of the model chosen for use. To be representative here, the estimates are taken based upon the popular TI MSP430 family, although possibly lower power dedicated components may be available.

In addition, the complexity of the multiplication operation depends on the size of the matrix used. In general for an $N \times N$ matrix it is an $O(N^3)$ process [20]. In the case for compressive sensing, however, where the **Φ** matrix is $M \times N$ this bound reduces to $O(M^{1.594}N)$ [21], significantly reducing the power required. Even so, based upon a **Φ** resolution of 16 bits, for any reasonable M and N it is likely that the MSP430 will have to be operated at the maximum clock frequency of 1 MHz, corresponding to an active mode power consumption of approximately 352 μW [22]. It is unlikely that portable EEG systems will be designed to have more than 64 channels, hence this power rating is considered to be the worst case scenario for systems having 64 or smaller number of channels.

Synchronization unit. For the purposes of the analysis here P_{Sync} and S are assumed to be negligible: they are essentially one time (start-up) operations that require the generation and transmission of approximately 48 bits (16 to initiate the random number process and 32 for the synchronization with the receiver). Compared to a continuous data rate in the range of kbps even for compressively sensed EEG this is deemed negligible.

Transmitter. The figure for J, the net energy per bit transmitted is simply taken from [23] which summaries the performance of several off-the-self transmitters finding that 50nJ/b is a conservative figure which should be readily achievable in most usage situations, and 5nJ/b is a more speculative figure for what may be possible. In this work this speculative 5nJ/b figure is used.

Table 3. Summary of the model parameters used and their justification

Parameter	Symbol	Value	Reasoning
Front end power	$P_{Amp} + P_{ADC}$	2 μW	From Tables 1 and 2.
Random number generator power	P_{RNG}	3 μW	From [19].
Matrix multiplication power	P_{DSP}	352 μW	From [22] and discussion above.
Seed and synchronization power	P_{Sync}	0 μW	From discussion above assumed negligible.
Transmitter energy required per bit transmitted	J	5nJ/b	From [28].
Net number of samples taken:	M	Variable	The effect of this will be investigated in Section 4.
Compressive sensing frame size:	N	750	Arbitrary choice to illustrate one performance point.
Nyquist sampling frequency	f_s	200 kS/s	From standard EEG specifications [13].
ADC sampling resolution	R	16 bits	Idealized value
Bits required to initialize random number process and synchronize with receiver	S	0 bit	From discussion above assumed negligible
Number of channels in the system	C	Variable	The effect of this will be investigated in Section 4.

4 Results and Discussion

Given the figures from Table 3 the implications of (6) and (7) can be investigated. Fig. 2 shows how the ratio M/N, which determines the amount of compression achieved as well as the end reconstruction error, affects the system power. In Fig. 2, N is arbitrarily set to 750 samples to limit the size of each matrix multiplication required. As may be expected, increasing M results in transmitting more data and so the system power consumption increases. The overall power consumption is also seen to be a strong function of the number of channels used.

This is illustrated more clearly in Fig. 3 which takes a compressive sensing operating point of M=80, N=750, and shows how the system power consumption varies with the number of channels present when compressive sensing is and isn't present. From this it is seen that for this operating point a compressive sensing based system is only feasible if more than 22 channels are to be present. When fewer channels than this are needed it is preferable to simply transmit the raw data.

This potentially has significant implications for augmented cognition applications of the EEG. For example, many augmented cognition applications such as [24] and [25] are using in the region of six channels. If this is sufficient for use there is no benefit to a compressive sensing based system, and optimizing the reconstruction performance and answering the open questions about basis functions and similar is not of practical interest at this time.

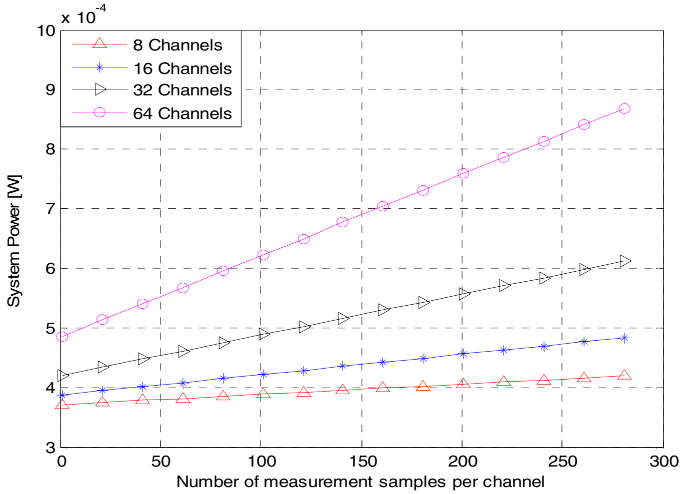


Fig. 2. The trade-off between the number of measurement samples taken (*M*), the number of channels used and the total system power consumption

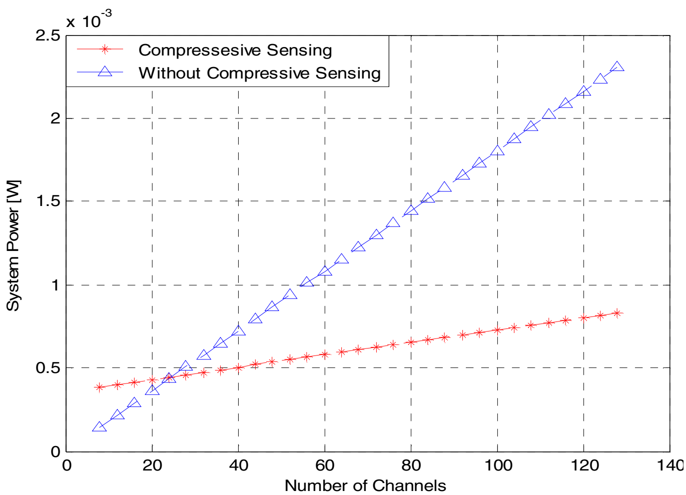


Fig. 3. The trade-off between the system power consumption and the number of channels (*C*) used for *M/N*=80/750 illustrates that a compressive sensing based system is only feasible when more than 22 channels are used

In contrast, there are other augmented cognition systems such as [26, 27] which are using 128 or more channels for recording. In this situation the use of compressive sensing is highly beneficial, with a reduction in system power consumption by 1.5mW being achievable for a 128 channel system. Using a conventional 30mWh small coin cell battery this could increase operational lifetime from 13 hours to 36 hours. In turn this can lead to significant improvements in the reliability, robustness and ease of use of systems allowing the accurate collection of physiological data.

5 Conclusions

Online data compression can be of significant use in facilitating the operation of portable EEG units from physically small batteries over a long period of time. In turn this aids the reliability and robustness of the overall system as the device is easier to use and more comfortable to wear. This paper has quantified the feasibility, from a power point of view, of using compressive sensing in order to provide this online data reduction.

Compressive sensing is a recent development in compression theory that states that it is possible to effectively sample a signal at a sub-Nyquist rate and yet still be able to accurately reconstruct the signal. Assuming that acceptable signal reconstruction is possible, this paper has presented a system modeling framework that quantifies the required power overhead for the compression system.

It was found that the feasibility of a compressive sensing based EEG system is a strong function of the number of channels present in the system; no benefit is present when less than 22 channels are needed (for the case considered here), but large power savings can be made when high numbers of channels are present. The feasibility of a compressive sensing based EEG system thus varies on an application-by-application basis, and the framework presented here can be used to assess this.

Given this result, there are potential benefits to using a compressive sensing system. Future work will thus focus on answering the many open questions still present: for example what basis functions and compression ratios can be used to minimize the reconstruction error.

References

1. Yates, D., Casson, A., Rodríguez-Villegas, E.: Low Power Technology for Wearable Cognition Systems. In: 13th International Conference on Human-Computer Interaction, pp. 127–136 (2007)
2. Aviyente, S.: Compressive sampling Framework for EEG Compression. In: IEEE/SP 14th Workshop on Statistical Signal Processing, pp. 181–184 (2007)
3. Lustig, M., Donoho, D., Santos, J., Pauly, J.: Compressive sampling MRI (A look at how CS can improve on current imaging techniques). IEEE Signal Processing Magazine 25(2), 72–82 (2008)
4. Donoho, D.: Compressive sampling. IEEE Trans. Inform. Theory 52(4), 1289–1306 (2006)
5. Candès, J., Wakin, M.: People hearing without listening: an introduction to compressive sampling. IEEE Signal Processing Magazine 25(2), 21–30 (2008)
6. Harrison, R., Charles, C.: A low-power low-noise CMOS amplifier for neural recording applications. IEEE Journal of Solid State Circuits 38(6), 958–965 (2003)
7. Yates, D., Villegas-Rodriguez, E.: An ultra low power low noise chopper amplifier for wireless EEG. In: The 49th IEEE Midwest Symposium on Circuits and Systems, pp. 449–452 (2006)
8. Menolfi, C., Huang, Q.: A Fully Integrated, Untrimmed CMOS instrumentation amplifier with sub microvolt offset. IEEE Journal of Solid State Circuits 34(3), 415–420 (1999)
9. Steyaert, M., Sansen, W., Zhongyuan, C.: A Micropower Low-noise monolithic instrumentation amplifier for medical purposes. IEEE Journal of Solid State Circuits 22(6), 1163–1168 (1987)

10. Wu, H., Xu, Y.: A Low-Voltage Low-Noise CMOS Instrumentation Amplifier for Portable Medical Monitoring Systems. In: IEEE-NEWCAS, pp. 295–298 (2005)
11. Yazicioglu, R.F., Merken, P., Puers, R., Van Hoof, C.: A 200μW Eight-Channel EEG Acquisition ASIC for Ambulatory EEG Systems. IEEE Journal of Solid State Circuits 43(12), 3025–3038 (2008)
12. Enz, C., Vittoz, E., Krummenacher, F.: A CMOS Chopper Amplifier. IEEE Journal of Solid State Circuits 22(3), 335–342 (1987)
13. Nuwer, M., Combi, G., Emerson, R., Fuglsang-Frederiksen, A., Guérit, J., Hinrichs, H., Ikeda, A., Luccas, F., Rappelsburger, P.: ICFN standards for digital recording of clinical EEG. Electroencephalography and Clinical Neurophysiology 106, 259–261 (1998)
14. Dlugosz, R., Gaudet, V., Iniewski, K.: Flexible Ultra Low Power Successive Approximation Analog-to-Digital Converter with Asynchronous Clock Generator. In: Canadian Conference on Electrical and Computer Engineering, pp. 1649–1652 (2007)
15. Verma, N., Chandrakasan, A.: An ultra low energy 12-bit rate-resolution scalable SAR ADC for wireless sensor nodes. IEEE Journal of Solid State Circuits 42(6), 1196–1205 (2007)
16. Bonfini, G., Garbossa, C., Saletti, R.: A switched Opam-based 10-b integrated ADC for ultra low power applications, VLSI-SOC (2003)
17. Gambini, S., Rabaey, J.: A 100KS/s 65dB DR $\Sigma - \Delta$ ADC with 0.65V supply voltage. In: 33rd European Solid State Circuits Conference, pp. 202–205 (2007)
18. López-Morillo, E., Gonzalez-Carvajal, R., Galan, J., Ramirez-Angulo, J., Lopez-Martin, A., Rodriguez-Villegas, E.: A Low-Voltage Low-Power QFG-based Sigma-Delta Modulator for Electroencephalogram Applications. In: IEEE BioCAS, pp. 118–121 (2006)
19. Holleman, J., Otis, B., Bridges, S., Mitros, A., Diorio, C.: A 2.92µW Hardware Random Number Generator. In: ESSCIRC Solid-State Circuits Conference, pp. 134--137 (2006)
20. Kakaradov, B.: Ultra-Fast Matrix Multiplication: An Empirical Analysis of Highly Optimized Vector Algorithms. Stanford Undergraduate Research Journal 3, 33–36 (2004)
21. Beling, P., Megiddo, N.: Using fast matrix multiplication to find basic solutions. Theoretical Computer Science 205(1-2), 307–316 (1998)
22. Texas Instruments: MSP430 DataSheet, `http://focus.ti.com/lit/ds/symlink/msp430f1101a.pdf`
23. XS110 UWB solution for media-rich wireless applications, `http://www.freescale.com/`
24. Kahol, K., Smith, M., Mayes, S., Deka, M., Deka, V., Ferrara, J., Panchanathan, S.: The Effect of Fatigue on Cognitive and Psychomotor Skills of Surgical Residents. In: 13th International Conference on Human-Computer Interaction, pp. 304–313 (2007)
25. Stevens, R., Galloway, T., Berka, C.: Exploring Neural Trajectories of Scientific Problem Solving Skill Acquisition. In: 13th International Conference on Human-Computer Interaction, pp. 400–408 (2007)
26. Huang, R.-S., Jung, T.-P., Makeig, S.: Event-Related Brain Dynamics in Continuous Sustained-Attention Tasks. In: 13th International Conference on Human-Computer Interaction, pp. 65–74 (2007)
27. Van Orden, K., Viirre, E., Kobus, D.: Augmenting Task-Centered Design with Operator State Assessment Technologies. In: 13th International Conference on Human-Computer Interaction, pp. 212–219 (2007)

Are You Really Looking? Finding the Answer through Fixation Patterns and EEG

Anne-Marie Brouwer[1], Maarten A. Hogervorst[1], Pawel Herman[2], and Frank Kooi[1]

[1] TNO Human Factors, Kampweg 5, 3769 ZG Soesterberg, The Netherlands
[2] Donders Institute for Brain, Cognition and Behaviour, The Netherlands
{anne-marie.brouwer,maarten.hogervorst,frank.kooi}@tno.nl, pawel.herman@donders.ru.nl

Abstract. Eye movement recordings do not tell us whether observers are 'really looking' or whether they are paying attention to something else than the visual environment. We want to determine whether an observer's main current occupation is visual or not by investigating fixation patterns and EEG. Subjects were presented with auditory and visual stimuli. In some conditions, they focused on the auditory information whereas in others they searched or judged the visual stimuli. Observers made more fixations that are less cluttered in the visual compared to the auditory tasks, and they were less variable in their average fixation location. Fixated features revealed which target the observers were looking for. Gaze was not attracted more by salient features when performing the auditory task. 8-12 Hz EEG oscillations recorded over the parieto-occipital regions were stronger during the auditory task than during visual search. Our results are directly relevant for monitoring surveillance workers.

1 Introduction

Imagine a man sitting behind a desk, looking at several monitors that display the images produced by surveillance cameras. His eyes are open and the radio is turned on. How do we know whether he is actually paying attention to what happens on the monitors rather than listening attentively to the radio or thinking about the groceries? We want to see whether we can distinguish between these kinds of situations - in particular, whether an observer's main current occupation is visual or not. We propose that fixation patterns and EEG could contribute to this distinction.

With respect to fixation patterns, we will test three different hypotheses.

Firstly, we expect that fixations will be guided more by bottom-up, low level visual features when observers are not occupied with a visual task compared to when they are. The idea is that if there is no explicit top-down, high level visual goal, low-level visual features will play a more important role. In order to test this hypothesis, we will compare fixation locations to saliency maps [1, 2]. Saliency maps indicate how saliency varies across the visual environment, based on low-level visual features such as, in our case, orientation, intensity and color. Rather than absolute values, center-surround values are used such that it is the contrast within a feature space that counts.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 329–338, 2009.

Secondly, we predict that in a visual search task fixations will be spread more than in an auditory task since in the first case, subjects have to scan an area whereas in the second case, they can choose to view one particular interesting area. We will both look at the standard deviation of fixation location and at measures defined by using Voronoi diagrams, following [3]. A Voronoi diagram is a division of an area into cells, with the borders of each cell surrounding a point of interest, in our case, a fixation. Every point within the cell is closer to the cell's fixation than to any other fixation. Small cell sizes go together with a densely fixated region whereas large cell sizes indicate occasional fixations. Fixation clutter produces a large difference in cell sizes, while evenly distributed fixations produce more similar cell sizes. Thus, fixation clutter can be quantified by the standard deviation of cell size or the skewness of the cell size distribution. [3] determined skewness in a free viewing condition, visual search in structured images and visual search in homogeneous images. Fixation clutter as determined by skewness decreased across these three conditions.

Thirdly, we expect that image patches around fixations will reveal features of a searched target (only when observers are performing a visual search task).

With respect to EEG correlates, we expect low alpha power over the parieto-occipital (visual) cortex when a visual search task is performed. Alpha refers to 8 to 12 Hz oscillatory EEG activity. It has been shown to be negatively correlated with visual (and not auditory) attention [4, 5] which is in agreement with the alpha inhibition hypothesis [6].

In order to determine whether a person is engaged in a visual task or not, the proposed fixation- and EEG cues could be combined. The particular (non)visual tasks included in this study are a visual search task, a visual judgment task and an auditory attention or short term memory task.

2 Methods

2.1 Equipment and Stimuli

The setup involved four computers. One was used to control the eye recordings, one to record EEG data, and one to present stimulus images. These computer units were synchronized. The fourth computer was used to present auditory stimuli.

Subjects were sitting in a shielded room (Faraday cage) with their head in a chin-rest and their right index and middle fingers on computer mouse buttons. Their eyes were approximately 65 cm away and at the height of the center of a 20 inch LCD monitor. Auditory stimuli were presented using speakers in the room. A Tobii eyetracker sampling eye position at 50 Hz, was positioned under the monitor.

EEG activity was recorded at the Fz, Cz, Pz, Oz, C3, C4, TP7, TP8, P3, P4, O1 and O2 electrode sites. A ground electrode was attached to the forehead and the reference was positioned on the left mastoid. The impedances between relevant pairs of electrodes were below 5 kΩ. The sampling rate was 256 Hz and the signal was bandpass filtered in the range of 0.1 - 100 Hz. Also, a 50 Hz notch filter was applied.

We used both natural and artificial images as visual stimuli (see for examples Figure 1). The natural images were pictures of landscapes, sometimes containing a (rather inconspicuous) military vehicle. These images were chosen from the Search_2 database [7] and from a larger collection of images of which Search_2 is a subset. The artificial images were generated by custom software, and consisted of a collection of colored simple geometric shapes. 20% of the images contained a green diamond and a blue pentagon.

The landscape images subtended the complete width of the monitor but not the complete height. The shapes images subtended the whole screen. Each image was presented for 5 s. In between images, a black screen was presented for 1 s.

The auditory stimuli consisted of the spoken letters a, b, c, x, y and z. Every 2.0 ± 0.2 s a letter randomly chosen from this sequence was presented for the entire duration of the experimental block.

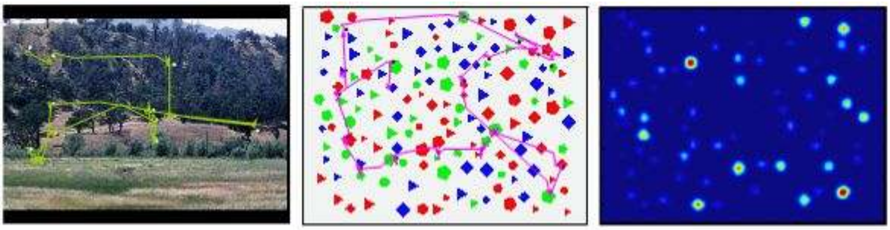


Fig. 1. On the left are examples of a landscape and a shapes image with accompanying gaze tracks and fixations as recorded during a trial of visual search. On the right is the saliency map from the shapes image example as computed by the Saliency Toolbox 2.1 [2].

2.2 Tasks

In the visual search task, subjects were asked to look for a target: a military vehicle in the landscape images or, either a green diamond (5 subjects) or a blue pentagon (6 subjects) in the shapes images. Subjects were requested to press the right mouse button when they had found the target and the left button if they could not find it. They were told to ignore the auditory presented letters.

In the visual judgment task, subjects were asked to judge the spread of the trees in the landscape images, and the spread of the red symbols in the shapes images. If subjects assessed these to be grouped, they pressed the left button, if they thought they were evenly spread, they pressed the right button. Subjects were told to ignore the auditory presented letters and advised to fixate their gaze at the center of the screen.

The auditory task was an auditory version of the Continuous Memory Task [8]. Subjects had to keep track of the letters 'a' and 'z'. Each time that one of these letters was presented for the first time, they were asked to right-click. When they were presented for the second time, subjects were supposed to left-click. After the second presentation subjects received feedback (a spoken 'correct' or 'false'). After the feedback, the counting started anew for the given letter. We refer to this task as an auditory task, but one could also consider it to be a short term memory or an auditory attention task. Although subjects were asked to watch the presented pictures, we stressed that keeping track of the letters was their main task.

2.3 Design and Procedure

10 naïve subjects plus the first author voluntarily participated in the study. Subjects were between 21 and 32 years old. Prior to the experiment they received written and spoken instructions.

The task order was random. For every task, a block of landscape images and a block of shapes images was presented (or vice versa). Each experimental block consisted of 40 images, shown in random order. Each block was preceded by a practice block of 5 images. Different images were used for each block. Before the start of each experimental block, a 9-point eye calibration session (ClearView 2.7) was performed to minimize fixation errors.

The first six subjects only performed the visual search and the auditory task; the last five subjects additionally performed the visual judgment task. Thus, the first group performed a total of 2 (tasks) * 2 (stimulus type) * 40 (number of repetitions) = 160 experimental trials, and the second group 3 (tasks) * 2 (stimulus type) * 40 (number of repetitions) = 240 experimental trials.

2.4 Analysis

We used the ClearView 2.7 default settings to define fixations. Examples of fixations plotted on the accompanying gaze track are presented in Figure 1. For each trial, we determined the number of fixations and the average fixation location. We also calculated the standard deviation of average fixation location between subjects. For trials with more than one fixation, the standard deviations in the horizontal and vertical direction were computed. In addition, we created Voronoi diagrams with every cell containing a fixation. The boundaries of the diagram were defined as the boundaries of the image. For each Voronoi diagram we normalized the cell sizes such that the average cell size was 1. Then we computed the standard deviation. A large standard deviation indicates more fixation clutter. Further, we computed the skewness of the distribution of cell sizes [3]. Higher values indicate more clutter.

For every stimulus image, we computed intensity, orientation, color and overall saliency maps according to [1] using the SaliencyToolbox 2.1 [2]. Figure 1 shows an example saliency map. For each fixation, we multiplied the saliency maps with a Gaussian window (standard deviation of 3 pixels) centered on the fixation. The average pixel value was determined. The results for all fixations within a trial were averaged. This value will be high when subjects fixate salient locations. To be able to compare between different images, the fixation saliency measure was defined as the ratio between this value and the average pixel value of the saliency map multiplied by the inverse of the Gaussian fixation windows (per trial, not per fixation).

For the shapes images, we wanted to identify the looked-for target from the fixations. Voronoi diagrams were calculated such that the center of each shape was the point around which cells were created. Next, we determined for each fixation location the color and the type of shape that was in the same cell as the fixation.

Repeated measures ANOVAs and paired t-tests were used to evaluate the results. For most variables, two ANOVAs were performed: one on the data of all 11 subjects with factors task ('visual search' and 'auditory') and stimulus type ('landscape' and 'shapes'), and one on the data of the last 5 subjects where the factor task had the levels 'visual judgment' and 'auditory'. Note that for the auditory task, the data in the second type of ANOVA (the grey symbols in the graphs) are a subset of the first (the black symbols). The significance level was set to 0.05. All significant effects are mentioned.

We visually inspected EEG signals in order to reject trials with excessive noise and artefacts. Then we performed power spectral analysis using the short-time Fourier transform (STFT) with Hanning window. We compared the average 8-12 Hz alpha power over the parieto-occipital region (Pz, P3, P4, Oz, O1 and O2) between the visual search and the auditory task, separately for shapes and landscapes images. The hypothesis that alpha power was larger in the auditory than in the visual task was verified using a one-sided Monte-Carlo testing procedure with dependent t-statistics. A Bonferroni correction was applied to correct for multiple comparisons (i.e. the significance level for each pair was set to 0.025). The entire EEG analysis was carried out using the Matlab package Fieldtrip (http://www.ru.nl/fcdonders/fieldtrip).

3 Results

3.1 Fixation Number

Figure 2 shows the average number of fixations per trial for each condition. In the visual task there are more fixations than in the auditory task, both when the task is visual search (effect of task $p<0.01$) and visual judgment ($p=0.02$). For visual search and the auditory task, there is an interaction between task and stimulus ($p<0.01$), such that the number of fixations is higher for the shape than the landscape stimuli in the search task, whereas there is no difference between the two in the auditory task.

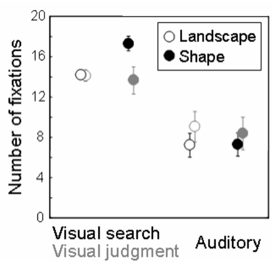


Fig. 2. Number of fixations. In all figures, error bars indicate standard errors of the mean. Black symbols represent data of all eleven subjects, grey symbols represent the subset of five subjects who also performed the visual judgment task.

3.2 Fixation Location

On average, subjects fixate around the center of the screen for all tasks and all stimuli. However, the average fixation locations are more variable between subjects in the auditory task compared to the visual tasks. The standard deviation of the average horizontal fixation location is 23 pixels (≈0.80 deg) for visual search, 28 pixels for visual judgment and 174 for the auditory task (all collapsed over stimulus type). For the vertical fixation location, these numbers are 22, 17 and 133, respectively. This suggests that in the auditory task, each individual subject had his/her own preferred fixation location in space.

3.3 Fixation Clutter

To investigate the spread of the fixation locations within trials, we computed the standard deviation of the horizontal and vertical fixation location for each trial that contained more than one fixation. Fixations were clearly more spread in the visual search condition as compared to the auditory condition (Figure 3, effect of task as indicated by p-values <0.01 for both the horizontal (3A, black symbols) and vertical direction (3B, black symbols)). There were also interactions with stimulus type (p=0.02 for the horizontal and p<0.01 for the vertical direction), indicating that the fixations were especially spread out when subjects were searching for vehicles compared to shapes, whereas the effect of stimulus type was the other way around during the auditory task. This indicates that the subjects scan more of the scene when looking for the target among shapes than in landscapes. For the subjects who performed the visual judgment task and the auditory task (Figure 3, grey values), there is only an effect of task (p=0.03 for the horizontal direction and p=0.02 for vertical), again indicating that the fixations are more spread in the visual condition compared to the auditory condition.

We also quantified clutter of fixations using Voronoi diagrams. Figure 4 displays the results of the two Voronoi measures, the standard deviation of the normalized cell size (A) and the skewness of the cell size distribution (B). The pattern of results is very similar for both measures: comparing visual search to the auditory task (black symbols) results in a main effect of task (p<0.01 for both standard deviation and skewness) and an interaction between task and stimulus type (p=0.04 for standard deviation and p<0.01 for skewness). The main effect of task indicates more fixation clutter in the auditory task than in visual search. The interaction suggests a trend toward more fixation clutter for the landscapes than for the shapes in the auditory condition, whereas it tends to be the other way around in the visual search condition. The ANOVAs performed on the data of the subjects who performed the visual judgment task did not result in any significant effect, neither for the standard deviation of the normalized cell size nor for the skewness.

Note that stimulus type can have an effect on fixation clutter in the auditory task. Paired t-tests comparing the two stimulus types in the auditory condition result in significant effects for the standard deviation of fixations in both directions, and for the standard deviation of the normalized Voronoi cell size (p-values <0.05). For skewness significance is approached (p=0.06).

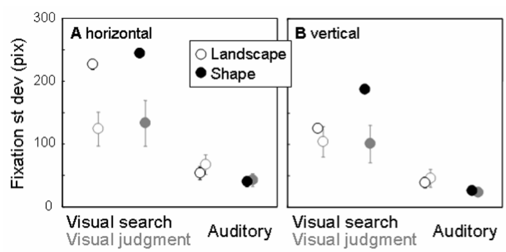


Fig. 3. Standard deviation of horizontal (A) and vertical (B) fixation locations

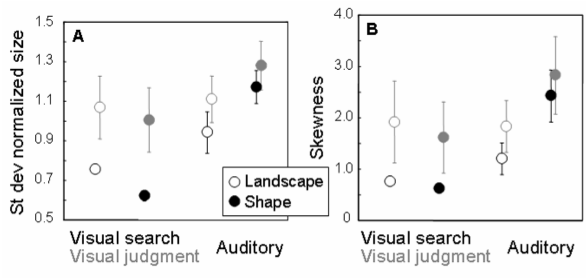


Fig. 4. Vonoroi measures of clutter: standard deviation of normalized cell size (A) and the skewness of cell size distribution (B)

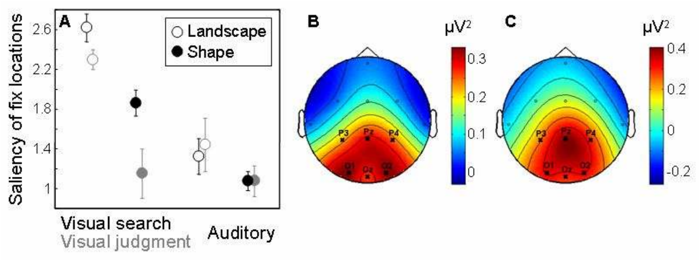


Fig. 5. Saliency of fixated locations (A) and distribution of the difference in alpha EEG power between the auditory and visual search tasks (B: landscapes, C: shapes)

3.4 Fixation Saliency

Figure 5 shows to what extent subjects fixate salient regions in the different conditions. When comparing the visual search task to the auditory task, the overall fixation saliency value is affected by task ($p<0.01$) and stimulus ($p<0.01$). Contrary to what we predicted, the task effect indicates that subjects fixate salient features more in the visual search task than in the auditory task. This is especially the case for the landscape stimuli. When comparing the visual search task to the auditory task, there is no significant effect of task, but an effect of stimulus type ($p=0.02$) and an interaction ($p=0.02$). When judging landscape rather than shape stimuli, subjects fixate highly salient features.

3.5 Fixation Target

For the shape stimuli, we had a closer look at the specific shapes that were fixated in the different conditions. Figure 6A shows the proportion of fixations close to blue, green and red shapes in the different tasks. In 6B, the data is split up by the form of the shape that subjects fixated. The data of the auditory and visual judgment task overlap in both graphs. This is in accordance with the saliency data where we did not observe any difference between the two tasks when the shapes were presented. One might have expected to find relatively many fixations close to red shapes in the visual judgment task (where the spread of red symbols had to be judged) but this is not

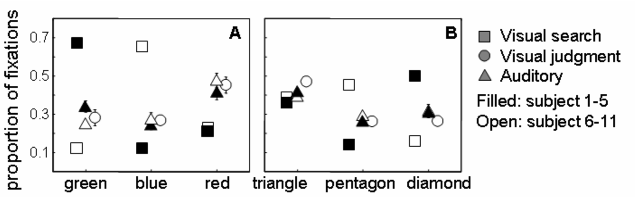


Fig. 6. Proportions of fixations close to differently colored (A) and formed (B) shapes. Note that all data points are together; except for the squares representing the search condition. The open squares indicate subject 1-5 who searched for the green diamond, the filled squares subject 6-11 who looked for the blue pentagon.

apparent. In general, red shapes (6A) and triangles (6B) seem to draw gaze. Our main objective was to determine the looked-for target from fixated features. Indeed, almost 70% of the fixations is on shapes of the same color as the target. Also, the form corresponding to that of the target is fixated most. Note that the target itself is hardly ever fixated because most trials did not contain a target, and if they did, those trials usually contained many more fixations on the other shapes.

3.6 EEG

As hypothesized, the average power of the parieto-occipital EEG alpha activity was higher during the auditory task than during visual search (p=0.015 for the landscape images with an average increase of alpha power of 52%, and p=0.022 for the shapes with an increase of 55%). Figure 5BC shows the distribution of the alpha power difference over the skull.

4 Discussion

Our experiment yielded several important findings.

Subjects make more fixations in both of the visual tasks compared to the auditory task. This is the case even though we asked subjects to view the pictures during the auditory task, and we tried to minimize the number of eye movements in the visual judgment task by instructing subjects to look at the center of the image. Note that the number of eye movements in visually judging the landscape images is the same as in searching them.

In the auditory task, subjects seem to have their own preferred fixation location in space. This suggests that fixation locations during non-visual tasks are not completely determined by low-level visual features. Although all subjects used a chin rest and an adjustable chair such that their eyes were approximately directed to the center of the screen, the preferred fixation location may have been influenced by the exact sitting and 'straight ahead' posture of the subject.

Fixations are more widely spread in both visual tasks than in the auditory task. The standard deviations of fixation locations indicate that when searching the shape stimuli, fixations were more spread than when searching the landscapes, while in the auditory task, fixations were less spread when viewing shape stimuli than landscapes.

Perhaps, the difficult shape stimuli forced subjects to scan a larger area when searching for the target, whereas, when there was no visual task to be done, subjects preferred scanning the landscape stimuli over the shape stimuli.

Computationally, the standard deviation does not depend on the number of data points. However, in fixation studies, the number of fixations could correlate positively with the standard deviation of fixation location: if subjects want to scan a larger area they will probably make more fixations. Still, this relation is not very strong as seen when Figures 2 and 3 are compared. This means that besides the number of fixations, fixation spread as indicated by the standard deviation may still add information when determining which task is being performed.

Fixation spread or clutter as measured using Voronoi diagrams provided similar results as when it was measured by the standard deviation. Fixations are less cluttered in the visual search task than in the auditory task, with an interaction effect such that landscapes cause more clutter in visual search than shape stimuli (consistent with [3]). The opposite effect of image type is observed in the auditory task. In contrast to the standard deviation, the Voronoi measures did not indicate any significant effect when comparing visual judgment to the auditory task. This is probably due to the larger variability in the Voronoi measures - the trend of the results is similar.

Interestingly, we found that the type of visual stimulus affects fixation spread in the auditory task. Thus, while the subject specific preferred fixation location in space during the auditory task indicates that fixation locations during non-visual tasks are not only determined by low-level visual features, visual stimuli do have an effect. This appears to be a higher level effect, such as an increased tendency to look around when natural, more meaningful stimuli are presented.

Our prediction that gaze is especially attracted by salient stimuli when subjects do not have a specific visual task did not bear out. On the contrary, especially for the landscape stimuli, subjects look more at salient features during visual tasks than during the auditory task. In the visual search task, this could be explained by the fact that some aspects of the target are salient. The findings suggest once again that during non-visual tasks low-level visual features do not determine fixation. We do not expect that these results can be easily reversed by using another definition of saliency since the significant effects of our variables indicate that the saliency maps as defined here are indeed meaningful.

We can deduce from fixations which target subjects are looking for, even if they hardly ever fixate the actual target. We showed this now for simple geometrical shapes, but if the eye tracking is accurate enough, it could work for more complex images as well. This could be done by applying different filters to the image patch around fixation and comparing the averaged results to randomly chosen patches.

The power of EEG alpha activity monitored over the parieto-occipital area is higher for the auditory task than visual search. This finding is completely in accordance with the alpha inhibition hypothesis [4, 5]. The emerging concept goes beyond the common notion of alpha as an ‘idling’ rhythm [9].

To sum up, our results suggest the following markers for a visual task: many fixations, a large fixation spread, and an average fixation location close to the center of the region of interest. Depending on the specific relevant items in the visual field, fixated features tend to be salient and can reveal which target the observer is looking for. The power of EEG alpha is low. Markers identified for a non-visual task are the following: few and cluttered fixations, average fixation locations far from the center of the region of interest, fixations that are not specifically directed at salient features and a relatively high EEG alpha power.

In order to effectively exploit these findings in monitoring surveillance workers, these different cues should be incorporated in a collective model. One way to address this challenge is to use classification techniques. This would entail training a classification model on features extracted from experimental data with the aim of enhancing the classifier's capability to generalize and categorize unseen data examples (in the given context, visual versus non-visual task). Pilot investigations into the robustness of this approach using solely the EEG data are promising.

Acknowledgements. BrainGain Smart Mix Programme of the Netherlands Ministry of Economic Affairs and the Netherlands Ministry of Education, Culture and Science. We also thank Sjaak Kriekaard, Antoon Wennemers, Carien Caljouw, Lex Toet and Jan van Erp.

References

1. Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. Vis. Res. 40, 1489–1506 (2000)
2. Walther, D., Koch, C.: Modeling attention to salient proto-objects. Neural Networks 19, 1395–1407 (2006)
3. Over, E.A.B., Hooge, I.T.C., Erkelens, C.J.: A quantitative measure for the uniformity of fixation density: The Voronoi method. Behav. Res. Meth. 38(2), 251–261 (2006)
4. van Dijk, H., Schoffelen, J.M., Oostenveld, R., Jensen, O.: Pre-stimulus oscillatory activity in the alpha band predicts visual discrimination ability. J. Neurosci. 28, 1816–1823 (2008)
5. Foxe, J.J., Simpson, G.V., Ahlfors, S.P.: Parieto-occipital 10 Hz activity reflects anticipatory state of visual attention mechanisms. Neuroreport 9, 3929–3933 (1998)
6. Klimesch, W., Sauseng, P., Hanslmayr, S.: EEG alpha oscillations: the inhibition-timing hypothesis. Brain Res. Rev. 53, 63–88 (2007)
7. Toet, A., Bijl, P., Valeton, J.M.: Image dataset for testing search and detection models. Optical Engineering 40(9), 1760–1767 (2001)
8. Veltman, J.A., Gaillard, A.W.K.: Physiological workload reactions to increasing levels of task difficulty. Ergonomics 5, 656–669 (1998)
9. Pfurtscheller, G., Stancak Jr., A., Neuper, C.: Event-related synchronization (ERS) in the alpha band: an electrophysiological correlate of cortical idling (review). Int. J. Psychophysiol. 24, 39–46 (1996)

"What Was He *Thinking*?": Using EEG Data to Facilitate the Interpretation of Performance Patterns

Gwendolyn E. Campbell[1], Christine L. Belz[1], and Phan Luu[2]

[1] Naval Air Warfare Center Training Systems Division, 12350 Research Parkway, Orlando, FL 32826

[2] Electrical Geodesics, Inc., Riverfront Research Park, 1600 Millrace Drive, Suite 307, Eugene, OR 97403

{Gwendolyn.Campbell,Christine.Belz}@navy.mil, pluu@egi.com

Abstract. Previous research has demonstrated that EEG data can be used to identify and remove unintentional responses from a data set (guesses and slips). This study sought to determine if removing this error variance has a significant impact on the interpretation of a trainee's performance. Participants were taught to recognize tank silhouettes. Multiple linear regression models were built for each participant based on three sets of their data: 1) all trials of their performance data, 2) only trials that were learned according to a state space analysis, and 3) their intentional data as identified by EEG. When compared to an expert model, each of the three models for every participant yielded a different diagnosis, indicating that filtering performance data with EEG data changes the interpretation of a participant's competence.

Keywords: electroencephalography, training, and student modeling.

1 Introduction

There is a growing movement to incorporate measures of physiological and neurological data collected from warfighters into military systems. The premise is that the systems could make intelligent adaptations on the basis of those measurements, in order to increase the overall effectiveness of the warfighter [1]. Training seems to be a particularly promising field for the incorporation of these data [2]. Many researchers in this area have proposed monitoring physiological data to ensure that the trainee is being kept at an optimal level of alertness and engagement during the exercise – neither bored nor overwhelmed [3].

Of course, alternative applications of neurophysiological data in training systems have been suggested as well. This work follows up on the proposal that electroencephalography data (EEG) could be used to support the process of diagnosing a trainee's underlying competence [4]. Currently, trainers make inferences about a trainee's competence based on the pattern of correct and incorrect actions that he or she takes during an exercise. While these data are obviously highly relevant, it has long been known that performance is not a perfect reflection of competence. Some actions, for example, represent guesses (lucky or unlucky) and slips (unintentional

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 339–347, 2009.

actions, which are more likely to occur when a person is working quickly) and thus are not representative of stable cognitive patterns.

In fact, it has been shown that EEG data can reliably discriminate between intentional and unintentional responses [5]. While these results are suggestive of the potential for neurophysiological data to support the accomplishment of training goals, many questions still remain. Of particular interest is the question of whether or not distinguishing between intentional and unintentional responses has any practical impact on the actual diagnosis of trainee performance. In other words, it has yet to be demonstrated that diagnosing only the intentional behaviors will lead a trainer to draw different conclusions about a trainee's underlying competence from what he or she would have concluded based on a diagnosis of the entire set of performance data.

We address this issue in the current study, using the Brunswik Lens Model [6] as our paradigm for diagnosing underlying trainee competence in a decision-making task. According to this paradigm, a mathematical model is derived relating trainee decisions to the characteristics of the environment or stimulus being processed. This model is interpreted as indicating which characteristics the trainee is using, and to what degree, when making an identification decision. An analogous model is built on either expert performance data or perfect performance data and the two models are compared. Discrepancies between the two models are interpreted as weaknesses in the student's strategy of using characteristics or cues to make decisions.

Our hypothesis is that using EEG data to remove guesses and slips (unintentional responses) from a trainee's performance data set will result in a different interpretation of that trainee's competence than would have been derived if the entire set of performance data had been modeled. As an additional control condition, we used a statistical technique to try to identify and remove guesses from each trainee's performance data set, to see if the EEG data had any impact over and above that which could be achieved by a simpler and cheaper methodology.

2 Method

2.1 Participants

Ten right-handed volunteers, 7 women and 3 men, over the age of 18 were recruited for this study. The mean age of the participants was 28 years (SD = 11; range: 18-48). Each received financial compensation for their participation.

2.2 Apparatus

A 256-channel HydroCel Geodesic Sensor Net (Electrical Geodesics, Inc., Eugene, OR) was used to acquire the EEG data. All recordings were referenced to Cz and all of the electrodes were kept below 70 KΩ. The EEG was bandpass filtered (0.1- to 100-Hz) and sampled with a 16-bit analog-to-digital converter at 250 s/s. Eprime© (Psychology Software Tools, Pittsburgh, PA) was used for stimulus control.

2.3 Materials

Participants were trained on identification of military vehicles in a computer-based learning program created in EPrime©. The images used were bitmap files scanned from images selected from the United States Marine Corps unclassified anti-armor training materials. They consisted of eight tank silhouettes: the ASU85, Centurion, Chieftain, Leopard, M60A1, T62, T72, and ZSU23-4. The program presented a bitmap file of the tank silhouette, allowed a fixed amount of time for a response, provided feedback, and kept a record of the stimuli presented as well as participant responses including reaction times. Trials were presented in a block randomized order so that for each consecutive eight trials each stimulus was presented once but in random sequence.

2.4 Procedure

After completing the informed consent paperwork, each participant was fitted with a 256-channel sensor array, and then began the computer-based learning task, which was divided into four stages. The first stage consisted of 120 trials divided into 15 randomized blocks of the eight tanks. The goal of this stage was to familiarize participants of the association between tank names and response keys. A target tank name was displayed in the center of the screen. Near the bottom of the screen a representation of the response keys with the tank names indicated on each key was displayed. The participant's task was to press the corresponding key as quickly as possible. Pressing the correct response key initiated the next trial in the program.

The goal of the second stage was to oblige participants to remember which response keys corresponded to each tank name. The task was identical to the task in the first stage except that the labeled keyboard display was removed. Participants engaged in seven randomized blocks of the eight tank names for a total of 56 trials. Feedback was given after each trial, and the next trial was initiated by a correct key press by the participant.

The third stage consisted of the primary learning task in which participants were asked to learn to identify the tank silhouettes. Each trial began with presentation of a silhouette in the center of the computer screen. Participants had 2000 milliseconds to identify the silhouette by pressing the appropriate response key. Immediately following the response, or if the response time ended before a key press was made, the participant was given feedback including information about the correctness of their response and the correct name of the tank. The feedback remained on the display for 2000 milliseconds, or until the participant pressed a key, upon which the computer screen went blank for 100 milliseconds and then the next trial began. This stage proceeded through 400 trials divided into 50 randomized blocks of the eight tank silhouettes.

Finally, the fourth stage consisted of the testing stage. Similar to the previous task, participants were asked to identify the tank silhouettes in a brief period of time (1000 milliseconds) by pressing the appropriate key. No feedback was provided, however, at any time during this stage. There were 32 test trials divided into four randomized blocks of the eight tank silhouettes.

Following these four stages participants filled out a standard questionnaire regarding the comfort of the 256-channel sensor array net and a debriefing questionnaire regarding the learning task. This was comprised of Likert scale ratings of both the difficulty of the learning task and usefulness of the feedback as well as questions in which participants were asked to describe the features that they used to identify the tank silhouettes and any learning strategies they may have used during the experiment.

3 Results

In preparation for modeling, the eight tank images were decomposed into a set of 7 features, for example, the ratio of the length of the gun barrel to the length of the vehicle body. This was accomplished partly through visual inspection by the authors and partly using subjective reports from pilot subjects. These features were sufficient to uniquely discriminate each of the images. The following paragraphs describe the steps used to build models. It should be noted that separate models were built for each participant and data were never combined across participants.

Each participant's data from stage three of the procedure (400 trials) were assembled into a table that contained one row per trial, and detailed the actual stimulus, the participant's response, the participant's reaction time and the values that those seven features took on for that stimulus. The analysis tool pack from Microsoft Excel ® was used to conduct a multiple linear regression on each participant's complete data set from stage three, with the constant set to zero, resulting in the first of the three models for each participant.

While logistic regression would technically have been more appropriate given the nature of the response (a vehicle name), a logical ordering of the vehicles (based on similarity) was imposed and a comparison of the two regression techniques indicated that they derived representations that were equally predictive of the participants' responses. The linear regression representation was used for this study because it provided cue weights that were easier to interpret within the context of the Brunswik Lens Model.

Next, following [7], state space analyses were applied separately to each participant's performance on each of the eight stimulus images, in an attempt to estimate the trial (if any) at which each image was reliably learned. Responses made before these learning points were discarded (as guesses) and the subset of performance data remaining was again analyzed by multiple linear regression.

The results of the state space analyses were also used as inputs to support the single trial analyses of the EEG data that was collected during stage three of the procedure. EEG-based indices were developed to discriminate between intentional (or learned) responses, guesses and slips. More details on the single trial analysis procedure that was applied can be found in [5]. Once the single trial analyses were complete, the results were used to identify and discard all of the responses that were not flagged as intentional and learned by the EEG signal. The remaining subset of data was used to generate the third model for each participant, following the procedure described above. Beta weights for each of the three models built for each participant can be found in Table 1.

In order to apply the Brunswik Lens Model paradigm, we needed to create one last model using hypothetical data from a "perfect participant." This was accomplished by using the correct stimulus images as the criterion in a regression analysis, instead of the responses given by a real participant. The beta weights and the 95% confidence intervals around those beta weights are presented in Table 2.

Table 1. Cue weights from Regression Equations Built from Three Different Subsets of Each Participant's Data

CUES	1	2	3	4	5	6	7
Participant #1							
All	0.16	3.07	0.77	-0.79	1.22	0.00	1.71
SSA	0.05	3.40	0.91	-0.93	1.21	0.00	1.89
EEG	-0.28	5.20	1.32	-1.40	1.74	-0.01	2.18
Participant #2							
All	-0.23	7.02	1.65	-1.55	3.02	-0.02	1.64
SSA	-0.37	7.17	1.70	-1.62	2.91	-0.02	1.83
EEG	-0.26	5.22	1.34	-1.39	1.75	-0.01	2.18
Participant #3							
All	0.60	2.76	0.36	-1.97	2.17	0.01	0.40
SSA	-1.02	0.00	-0.26	-0.38	-1.11	0.03	3.25
EEG	-0.21	4.98	1.26	-1.37	1.64	-0.01	2.17
Participant #4							
All	0.71	-0.67	0.28	-1.23	0.46	0.02	0.74
SSA	-0.11	-4.14	0.50	0.00	1.38	0.05	0.92
EEG	-0.27	4.95	1.29	-1.37	1.71	-0.01	2.15
Participant #5							
All	0.88	4.81	-0.57	-2.06	1.16	0.00	0.94
SSA	0.71	5.10	0.13	-1.64	0.59	-0.01	1.81
EEG	-0.30	5.09	1.32	-1.39	1.76	-0.01	2.17
Participant #6							
All	0.46	-1.11	-0.67	-1.46	-0.30	0.03	0.96
SSA	0.72	0.00	0.36	-0.73	0.00	0.01	1.64
EEG	-0.18	5.05	1.31	-1.31	1.59	-0.01	2.22
Participant #7							
All	-0.14	4.58	1.13	-1.33	1.75	-0.01	1.97
SSA	-0.11	3.92	1.18	-1.09	1.44	0.00	2.09
EEG	-0.22	4.80	1.37	-1.24	1.64	-0.01	2.19
Participant #8							
All	0.44	5.60	-0.38	-2.89	1.56	0.00	1.14
SSA	-0.07	4.92	1.04	-1.72	2.00	-0.01	1.71
EEG	-0.36	5.26	1.42	-1.38	1.88	-0.01	2.18
Participant #9							
All	0.06	-0.46	-1.18	-0.98	-0.19	0.04	1.10
SSA	0.24	0.17	-0.45	-1.07	-0.19	0.02	1.59
EEG	-0.27	4.94	1.32	-1.33	1.69	-0.01	2.18
Participant #10							
All	0.88	3.32	-0.05	-1.38	0.31	0.00	1.55
SSA	0.85	0.00	0.14	-0.47	-0.46	0.01	1.87
EEG	-0.38	5.08	1.43	-1.34	1.89	-0.01	2.16

Table 2. Beta weights and 95% confidence intervals in hypothetical perfect participant model

CUES	1	2	3	4	5	6	7
Perfect Performance Model							
Lower 95th confidence interval	-0.37	4.43	1.20	-1.44	1.43	-0.01	2.14
Beta weights	-0.25	4.93	1.38	-1.27	1.70	-0.01	2.19
Upper 95th confidence interval	-0.12	5.43	1.56	-1.11	1.97	-0.01	2.24

Table 3. Summary of beta weight comparisons between the perfect participant's model and each of the three models built per participant

	# of Beta Weights	
	β > Upper Bound OR β < Lower Bound on P.P. Model	Lower < β < Upper Bounds of P.P. Model
Participant 1		
All	7	0
SSA	7	0
EEG	0	7
Participant 2		
All	6	1
SSA	6	1
EEG	0	7
Participant 3		
All	7	0
SSA	7	0
EEG	0	7
Participant 4		
All	6	1
SSA	7	0
EEG	0	7
Participant 5		
All	6	1
SSA	6	1
EEG	0	7
Participant 6		
All	7	0
SSA	7	0
EEG	0	7
Participant 7		
All	3	4
SSA	6	1
EEG	0	7
Participant 8		
All	6	1
SSA	6	1
EEG	0	7
Participant 9		
All	7	0
SSA	7	0
EEG	0	7
Participant 10		
All	6	1
SSA	7	0
EEG	1	6

The final step was to conduct the diagnosis of each participant model. This was accomplished by comparing each of the beta weights in the participant's model to the 95% confidence intervals around the beta weights in the perfect model. If a beta weight fell outside of a confidence interval, the qualitative interpretation would be that the participant did not use information available in that cue appropriately.

For example, consider the regression equation built for participant #1 using all of his or her data. The beta weight on the second cue is 3.07. The perfect participant model shows a beta weight of 4.93, and the lower bound on the 95% confidence interval around this beta weight is 4.43. This could be interpreted as saying that, according to the participant's data, the participant is under-utilizing the information available in this cue or characteristic of the vehicle images.

Next, consider the same cue for participant #2. This participant has a beta weight of 7.02 on the second cue, which is above the upper bound of 5.43 on the 95% confidence interval around the beta weight in the perfect participant's model. It would appear that participant #2 over-relies upon information contained in this feature of the vehicle images when making his or her identification decision.

The results of this comparison are summarized in Table 3. The comparison of interest is, for each participant, the interpretation of the accuracy of his or her cue usage to make identification decisions within each of the three models. More specifically, an examination of the table reveals a high degree of overlap in the first two rows of each participant's section of the table, and a large deviation in the third row of each participant's section of the table. In other words, the model based on EEG information led to a different diagnosis of competence for every single one of the 10 participants.

4 Discussion

Currently, the measurement and analysis of electroencephalographic (EEG) signals can be a complicated, cumbersome and costly procedure. While there has been some scientific work suggesting that single trial analysis of EEG data can distinguish between intentional and unintentional responses given in a training context [5], that is only a first step towards determining the potential practical value added of using neurophysiological data in a real training setting. In this paper, we investigated the question of whether or not making this discrimination at a response-by-response level has the potential to influence a more global diagnosis of a trainee's cognitive strategy. If discriminating intentional from unintentional responses doesn't change the ultimate diagnosis of a trainee's strengths and weaknesses, then it is unlikely that this extra step is worth the required resources.

We used multiple linear regression equations to estimate the extent to which a trainee was over-, under- or appropriately using the various features of tanks to help identify them. We evaluated the appropriateness of their cue usage by comparing cue weights from their strategy regression equations to cue weights from the strategy regression equation of a hypothetical perfect participant. More specifically, we concluded that a participant was appropriately using a particular vehicle feature if his cue weight on that feature fell within the 95% confidence intervals around the perfect participant's cue weight for that feature.

In total, we built three equations for each participant. We built the first equation using all of that participant's data. Next, we applied a statistical technique to try to discriminate guesses from learned responses. The second equation for each subject was based on only the subset of responses that appeared to be learned according to this state-space analysis. Finally, we used single trial analyses of EEG data to discriminate between guesses, slips and intentional responses. The third equation for each subject was based on only the subset of responses that appeared to be intentional according to this neurophysiological analysis.

As our results clearly demonstrate, when compared to a diagnosis based on the entire set of responses given by a single participant, the statistical technique of identifying learned responses had little impact on the conclusions that a trainer would draw about the trainee's mental strategy. However, using the EEG-based filter to identify intentional responses had a dramatic impact on the conclusions that a trainer would draw for every single one of our ten participants. In each case, the diagnosis would flip from indicating that the participant was using few, if any, cues appropriately to indicating that the participant was using most, if not all, cues appropriately. Needless to say, these two sets of diagnoses would lead to very different instructional "next steps" for these trainees.

The fact that the use of the EEG filter led to the conclusion that most of the trainees were using all of the available information appropriately to make their identifications is not really surprising, given our training methodology. Remember that trainees were given the correct identification of each vehicle after every response. What this suggests is that this particular training paradigm led to accurate learning and that the trainees may have been further along that learning path than their performance data alone would lead us to believe.

While this work moves us one step closer to addressing the practical question of whether or not the incorporation of EEG-based measurement in a training system has value added, it is still not a final answer. We have demonstrated that the use of an EEG-based filter may lead to a different diagnosis of a trainee's underlying competence, however these data do not tell us if that diagnosis is, in fact, more accurate. The next step, which we are currently working on, is to see if using the EEG-based diagnosis to control the instructional response is either more effective or more efficient than relying on the trainee's entire data set.

It should be noted that there are also technical challenges that must be overcome, even if the EEG-based diagnosis does turn out to be more accurate. The method we used to conduct the single trial analysis of EEG data was largely data-driven and reasonably time consuming. To truly have practical application in a military training system, the EEG analyses would need to be automated and able to run in very-close to real-time.

Finally, of course, the fact that the EEG-based diagnoses differed substantially from the full data diagnoses for this particular training context does not guarantee that it will always have an impact. There could easily be many environments in which the use of this technology does not confer any advantage. For example, in slow moving domains that allow for a lot of deliberation before taking a single action, we would not expect to see a lot of unintentional responses that needed to be filtered out with EEG data. Similarly, in a domain that allowed operators to "undo" an accidental action, the use of EEG data to identify these slips would be overkill. Also, statistical techniques to distinguish intentional (or reliable) from unintentional (or unreliable) data are likely to be more cost effective then neurological data when there is the opportunity to collect a large enough sample of performance data from a trainee.

Despite the limitations of this study and the possible limitations on the use of this technology, we think that this work represents an important step forward towards the goal of effectively incorporating neurophysiological measurement into the assessment and diagnosis of trainee performance patterns. Our data have shown that, at least under some circumstances, the use of EEG data to filter the corresponding set of performance data can have a substantial impact on the conclusions that a trainer would draw about the trainee’s underlying knowledge and competence.

References

1. Berka, C., Levendowski, D., Ramsey, C., Davis, G., Lumicao, M., Stanney, K., Reeves, L., Regli, S., Tremoulet, P., Stibler, K.: Biomonitoring for Physiological and Cognitive Performance during Military Operations. In: Caldwell, J.A., Wesensten, N.J. (eds.) Proceedings of the SPIE, vol. 5797, pp. 90–99 (2005)
2. Dickson, B., Belyavin, A.: The use of electrophysiological markers of expertise to configure adaptive training systems. In: Schmorrow, D., Nicholson, D., Drexler, J., Reeves, L. (eds.) Foundations of Augmented Cognition, 4th edn., pp. 138–144. Strategic Analysis, Inc. and the Augmented Cognition International Society, Arlington (2007)
3. DuRousseau, D.R., Mannucci, M.A., Stanley, J.P.: Will augmented cognition improve training results? In: Schmorrow, D. (ed.) Foundations of Augmented Cognition, vol. 2, pp. 956–963. Lawrence Erlbaum & Associates Inc., Mahwah (2005)
4. Luu, P., Campbell, G.E.: “Oops, I did it again”: Using neurophysiological indicators to distinguish slips from mistakes in simulation-based training systems. In: Schmorrow, D. (ed.) Foundations of Augmented Cognition, vol. 2, pp. 941–945. Lawrence Erlbaum Associates Publishers, Mahwah (2005)
5. Campbell, G.E., Luu, P.: A preliminary comparison of statistical and neurophysiological techniques to assess the reliability of performance data. In: Schmorrow, D., Nicholson, D., Drexler, J., Reeves, L. (eds.) Foundations of Augmented Cognition, 4th edn., pp. 119–127. Strategic Analysis, Inc., Arlington (2007)
6. Campbell, G.E., Buff, W.L., Bolton, A.E.: Viewing training through a fuzzy lens. In: Kirlik, A. (ed.) Adaptation in Human-Technology Interaction: Methods, Models and Measures, pp. 149–162. Oxford University Press, Oxford (2006)
7. Smith, A.C., Frank, L.M., Wirth, S., Yanike, M., Hu, D., Kubota, Y., Graybiel, A.M., Suzuki, W.A., Brown, E.N.: Dynamic analysis of learning in behavioral experiments. The Journal of Neuroscience 24(2), 447–461 (2004)

Motion-Sickness Related Brain Areas and EEG Power Activates

Yu-Chieh Chen[1], Jeng-Ren Duann[1,2], Chun-Ling Lin[1], Shang-Wen Chuang[1], Tzyy-Ping Jung[1,2], and Chin-Teng Lin[1]

[1] Brain Research Center, National Chiao-Tung University, Hsinchu, 30010, Taiwan
[2] Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California San Diego, La Jolla, CA, 92093-0961 USA
mandychen.ece92g@nctu.edu.tw, duann@sccn.ucsd.edu, geralate.ece95g@nctu.edu.tw, swchuang@gmail.com, jung@sccn.ucsd.edu, ctlin@mail.nctu.edu.tw

Abstract. This study investigates electroencephalographic (EEG) correlates of motion sickness in a virtual-reality based driving simulator. The driving simulator comprised an actual automobile mounted on a Stewart motion platform with six degrees of freedom, providing both visual and vestibular stimulations to induce motion-sickness in a manner that is close to that in daily life. EEG data were acquired at a sampling rate of 500 Hz using a 32-channel EEG system. The acquired EEG signals were analyzed using independent component analysis (ICA) and time-frequency analysis to assess EEG correlates of motion sickness. Subject's degree of motion-sickness was simultaneously and continuously reported using an onsite joystick, providing non-stop psychophysical references to the recorded EEG changes. Five Motion-sickness related brain processes with equivalent dipoles located in the left motor, the parietal, the right motor, the occipital and the occipital midline areas were consistently identified across all subjects. These components exhibited distinct spectral suppressions or augmentation in motion sickness. The results of this study could lead to a practical human-machine interface for noninvasive monitoring of motion sickness of drivers or passengers in real-world environments.

Keywords: EEG, ICA, motion-sickness, delta, theta, alpha, time-frequency.

1 Introduction

Motion-sickness is a common experience to everybody, and it has provoked a great deal of attentiveness in plenty of studies. The sensory conflict theory that came about in the 1970's has become the most widely accepted theorem of motion-sickness among scientists [1]. The theory proposed that the conflict between the incoming sensory inputs could induce motion-sickness. Accordingly, new research studies have appeared to tackle the issue of the vestibular function in central nervous system (CNS). In the previous human subject studies, researchers attempt to confirm the brain areas involved in the conflict in multi-modal sensory systems by means of clinical or anatomical methods. Brandt et al. demonstrated that the posterior insula in

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 348–354, 2009.

human brain was homologous to PIVC in the monkey by evaluating vestibular functions in patients with vestibular cortex lesions [2]. In agreement with previous clinical studies, the cortical activations during caloric [3] and galvanic vestibular stimulation [4] had been studied by functional imaging technologies such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI). To overcome the temporal limitation of the two imaging modalities, some studies have investigated the vestibular information transmission in time domain. monitoring the brain dynamics induced by motion-sickness because of its high temporal resolution and portability De Waele et al. , for example, applied current pulse stimulation to patients' vestibular nerve to generate vestibular evoked potentials [5]. In the study, thirty active scalp electrodes were used to record evoked potentials. By means of dipole imaging method, five distinct cortical areas were modeled from the recorded scalp signals. However, there is no general consensus on the motion-sickness related brain areas among the previous studies.

The EEG studies related to motion-sickness can be divided into two groups according to the types of stimuli: vestibular and visual. Vestibular stimuli were normally provided to the subjects with rotating chair [6-7], parallel swing [8], and cross-coupled angular stimulation [9] to induce motion-sickness. Theta power increases in the frontal and central areas were reported to be associated with motion-sickness induced by parallel swing [8] and rotating drum [6-7]. Chelen et al. [9] employed cross-coupled angular stimulation to induce motion-sickness symptoms and found increased delta- and theta-band power during sickness but no significant change in alpha power. Visually induced motion-sickness is also commonly studied in previous studies. Visually induced sickness can be provoked with an optokinetic drum rotating around the yaw axis. This situation can cause a compelling sense of self-motion (called vection). Vestibular cues indicate that the body is stationary, whereas visual cues report the body is moving. Hu et al. investigated MS triggered by the viewing of an optokinetic rotating drum and found a higher net percentage increase in EEG power in the 0.5-4 Hz band at electrode sites C3 and C4 than in the baseline spectra. [10]. This study employees ahe driving simulator comprised an actual automobile mounted on a Stewart motion platform with six degrees of freedom, providing both visual and vestibular stimulations to induce motion-sickness and accompanied EEG dynamics.

2 Materials and Methods

2.1 Experimental Paradigm

Unlike the previous studies, we provided both visual and vestibular stimuli to participant through a compelling VR environment consisting of 360^{o} projection of VR scene and a motion platform with six degree-of-freedom to induce motion-sickness. With such a setup, we expected to create motion-sickness in a manner that is close to that in daily life.. During the experiment, the subjects were asked to sit inside an actual vehicle mounted on a motion platform, with their hands holding a joystick to report their sickness level continuously. The VR scenes simulating driving in a tunnel were programmed to eliminate any possible visual distracter and shorten the depth of visual field such that motion-sickness could be easily induced. A three-section experimental protocol (shown in Fig. 1) was designed to induce motion-sickness.

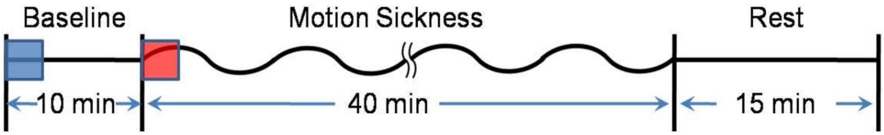


Fig. 1. Experimental design of motion-sickness experiments

First, the baseline section contained a 10-minute straight road to record the subjects' baseline state. Then, a 40-minute motion-sickness section, which consisted of a long winding road, was presented to the subjects to induce motion-sickness. Finally a 15-minute rest section with a straight-road condition was displayed for the subjects to recover from their sickness. The level of sickness was continuously reported by the subjects using a joystick with continuous scale on its side. The experimental setting successfully induced motion sickness to more than 80% of subjects in this study.

2.2 Subjects

Twenty-four healthy, right-handed volunteers (15 males and 9 females, ages from 21 to 24 year-old with an average of 22.1 year-old) with no history of gastrointestinal, cardiovascular, or vestibular disorders, no drug or alcohol abuse, no current medication, and having normal or corrected-to-normal vision participated in this experiment. Among the 24 participants, four having no motion-sickness at all and one being too sick were excluded from further data analysis. Therefore, the EEG data and subjective MS level from 19 subjects were included in the further data analysis.

2.3 Data Acquisition and Analysis

Thirty two-channel EEG signals were acquired at the sampling rate of 500 Hz using NuAmps (BioLink Ltd., Australia). The acquired multi-channel EEG signals were first down sampled to 250 Hz. A high pass filter with a cut-off frequency at 1 Hz with transition band of 0.2 Hz was used to remove baseline-drifting artifacts. Then, a low-pass filter with cut-off frequency at 50 Hz with transition band of 7 Hz was applied to the signal to remove muscle artifacts and line noise. During the experiment, the sickness level was continuously reported by subject using a joystick with continuous scale, which was synchronized with the EEG signals. The sickness level was ranged from 0 to 5. The continuous sickness level instead of the traditional motion-sickness questionnaire (MSQ) [11] used in this study gave us real-time sickness level ratings without interrupting the experiment for the subjects to fill out the questionnaire. The new rating system provides continuous information of sickness level and also ensures the quality of EEG signals. A moving 100s window was used to smooth the continuous subjective sickness ratings using steps of 1s.

The acquired EEG signals were analyzed using independent component analysis (ICA) and time-frequency analysis to assess the involved brain regions/circuits during motion-sickness. Then, components with similar scalp topographies, dipole locations and power spectra from many subjects were grouped into component clusters [12]. The activations of independent components (ICs) were then correlated to the MS level to investigate the changes before, during and after motion-sickness sessions.

3 Results

3.1 Single-Subject Time-Frequency Response

The EEG dynamics related to the motion-sickness level were investigated by applying time-frequency analysis to the ICs in the five selected ICs clusters. Fig. 2 shows the time-frequency response of three ICs, the occipital, parietal and right motor components, from one of the 19 subjects. Among these ICs, the alpha power of the left occipital component (Fig. 2B) increased as the MS level increased. In addition, the alpha power changes in parietal and motor areas are also partially co-varied with MS level as shown in Fig. 2C and 2D. As a result, these three IC clusters might be most related to motion sickness as revealed by their time-frequency dynamics.

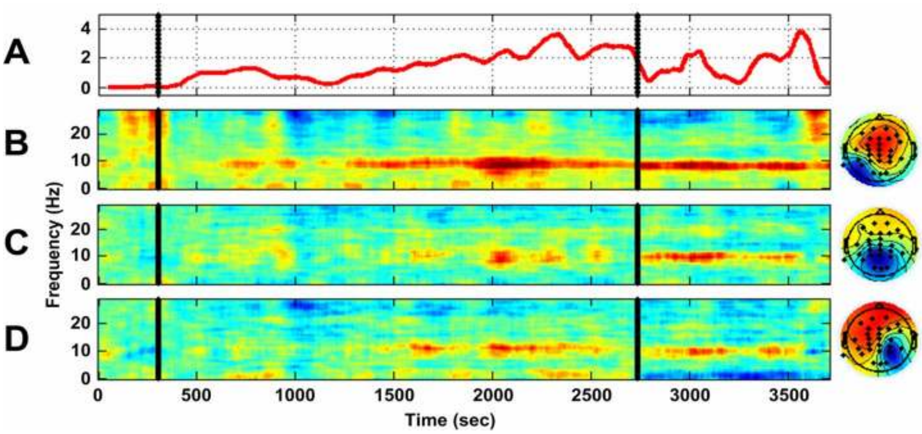


Fig. 2. Single-subject time-frequency results in occipital, parietal and right motor components. The recorded sickness level was shown as red curve in A. Three ICA components were selected as MS-related components in which their time-frequency responses were changed with sickness level. ICA power increases with the severity of motion-sickness were observed in alpha band.

3.2 Cross-Subject EEG Activities Related to MS

Brain signals can be sensitive to any environmental changes. Thus, the EEG signals acquired under different conditions may be confounded by different experimental variables. For example, when the experiment entered the winding-road section, the car began to sway left and right with the VR scene of the curved road, providing both visual and body sensation stimuli to the subjects which might induce large brain responses. As a result, simply comparing the EEG power associated with motion-sick and non-motion sick conditions may fail to dissociate such confounding effects from the main motion-sickness ones. This baseline difference among EEG power spectra associated with the different road conditions must be considered when the MS-related EEG power changes are evaluated. Therefore, the EEG power spectral changes in three periods were initially examined: (1) baseline - the first 3 minutes of the baseline straight road section, (2) low MS level - the first 3 minutes of the curved road section,

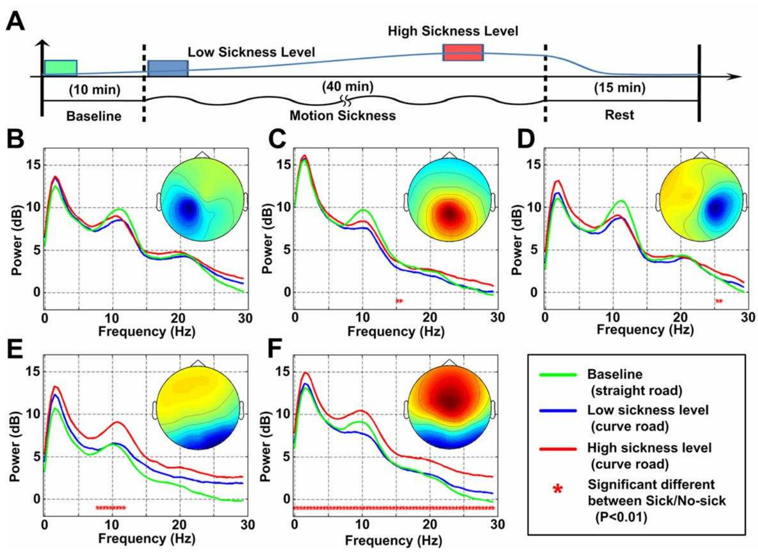


Fig. 3. Statistical comparison of IC power spectra under different conditions. The averaged ICA power spectra of baseline (straight-road / no-sick) were plotted as green lines, the power spectra of low sickness level (curve-road/ no-sick) were shown in blue, and the power spectra of low sickness level (curve-road/ sick) were shown in red. The red stars indicated significant different between sick and no-sick. A paired-sample Wilcoxon signed rank test was applied to the ICA power spectra to verify the significance of the power difference. The significant level was set at p <0.01 in this research.

and (3) high MS level - the first 3 minutes after the highest sickness rating. The power spectra in these three time periods (baseline, low-sickness and high-sickness) were then averaged among subjects in each selected IC cluster. A paired Wilcoxon signed rank test was performed on the averaged ICA power spectra to evaluate the statistical significance of relationship between the power difference and both the road conditions and the motion-sickness level.

Figure 3 compares the mean component power spectra of the IC clusters for different MS levels or under various road conditions. The EEG spectral difference associated with the different road conditions can be assessed by comparing the baseline power spectra (green traces in Figs. 3B-F) and the low-MS level spectra (blue traces). Evidently, the alpha power of the right, left motor and the parietal components were suppressed from the straight-road driving to the winding-road driving as the car swayed from side to side. Additionally, significant alpha power suppression in the occipital midline IC cluster was also observed (Fig. 3F). Comparing the component power spectra under low MS level (blue traces in Fig. 3) and high MS level (red traces) revealed MS-related spectra changes. The red asterisks in Fig. 3 indicate the frequency bins where the component EEG power differed significantly between the maximum and minimum sickness levels under the same curve road condition (p<0.01, Wilcoxon signed rank test). The alpha power of the occipital IC cluster (Fig. 3E) increased significantly with the MS level, whereas the occipital midline component cluster exhibited broadband spectral elevation at high MS.

3.3 Correlation between MS Level and EEG Responses

Figure 4 shows the overall correlations between the component spectra and their corresponding MS levels of the five clusters. The correlation coefficients in the alpha band exceed those in other frequency bands in all five clusters. The maximum correlation coefficient in the alpha band is 0.5 in the occipital midline components, while the correlation coefficients are approximately 0.4 in other IC clusters.

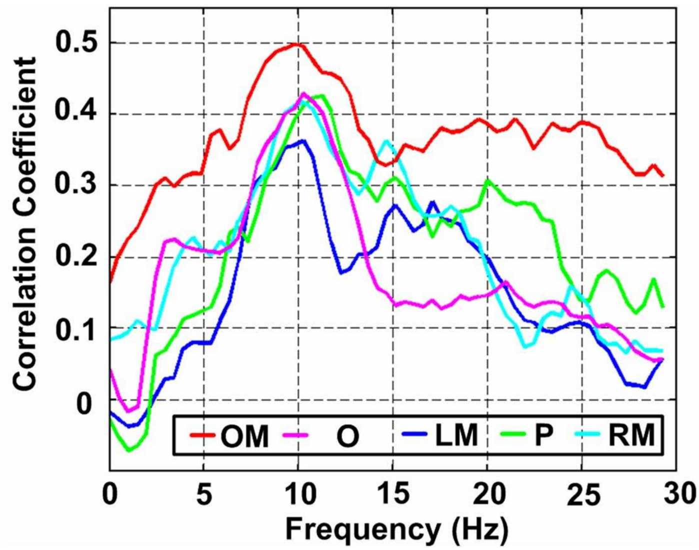


Fig. 4. Correlation between sickness level and the EEG spectra of the five component clusters

4 Discussion

The brain activities related to motion-sickness were studied here by using the technologies of EEG and the dynamic motion platform. ICA was used to separate distinct motion-sickness related EEG processes in the bilateral motor, parietal, occipital and occipital midline regions.

Alpha power suppressions correlated to motion stimuli (Fig.3 B-D) were found in parietal and the two somatosensory areas (the right and left motor areas). The power suppression can be referred to the blocking or desynchronization of central mu rhythms. This suggests that these brain areas might be influenced by vestibular inputs during the experiments. However, the frequency responses in these three brain areas were also affected by severity of motion-sickness. In Fig. 3E-F, we found alpha power increases with sickness-level, especially in the parietal lobe. The results are consistent with a gravity experiment proposed by [13]. They showed that the 10-Hz oscillations in the parieto-occipital and sensorimotor areas increased in the absence of gravity. They also suggested that since parietal lobe is situated at the conjunction between the visual and the sensorimotor cortex and thus it might be involved in integrating the multi-modal sensorimotor inputs.

Correlation analysis (Fig. 4) suggested that the power responses in the occipital midline components were highly correlated with subjective sickness-level, comparing to other brain areas. It suggests that the activations in the occipital midline, followed after sensorimotor integration in parietal and motor areas, might be used as a countermeasure to evaluate self-conscious of motion-sickness.

References

1. Reason, J.T., Brand, J.J.: Motion-sickness. Academic Press, London (1975)
2. Brandt, T., Dieterich, M., Danek, A.: Vestibular cortex lesions affect the perception of verticality. Annals of neurology 35(4), 403–412 (1994)
3. Fasold, O., von Brevern, M., Kuhberg, M., Ploner, C.J., Villringer, A., Lempert, T., Wenzel, R.: Human vestibular cortex as identified with caloric stimulation in functional magnetic resonance imaging. NeuroImage 17(3), 1384–1393 (2002)
4. Lobel, E., Kleine, J.F., Le Bihan, D., Leroy-Willig, A., Berthoz, A.: Functional MRI of galvanic vestibular stimulation. The journal of neurophysiology 80(5), 2699–2709 (1998)
5. De Waele, C., Baudonniere, P.M., Lepecq, J.C., Tran Ba Huy, P., Vidal, P.P.: Vestibular projections in human cortex. Experimental Brain Research 141, 541–551 (2001)
6. Wood, C.D., Stewart, J.J., Wood, M.J., Struve, F.A., Straumanis, J.J., Mims, M.E., Patrick, G.Y.: Habituation and motion-sickness. Journal of clinical pharmacology 34, 628–634 (1994)
7. Wood, S.J.: Human otolith-ocular reflexes during off-vertical axis rotation: effect of frequency on tilt-translation ambiguity and motion-sickness. Neuroscience letters 323(1), 41–44 (2002)
8. Wu, J.P.: EEG changes in man during motion-sickness induced by parallel swing. Space Medicine and Medical Engineering 5(3), 200–205 (1992)
9. Chelen, W.E., Kabrisky, M., Rogers, S.K.: Spectral analysis of the electroencephalographic response to motion-sickness. Aviation, space, and environmental medicine 64(1), 24–29 (1993)
10. Hu, S., McChesney, K.A., Player, K.A., Bahl, A.M., Buchanan, J.B., Scozzafava, J.E.: Systematic investigation of physiological correlates of motion-sickness induced by viewing an optokinetic rotating drum. Aviation, space, and environmental medicine 70(8), 759–765 (1999)
11. Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G.: Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. International Journal of Aviation Psychology 3(3), 203–220 (1993)
12. Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., Sejnowski, T.J.: Analysis and Visualization of Single-Trial Event-Related Potentials. Human Brain Mapping 14(3), 166–185 (2001)
13. Cheron, G., Leroy, A., De Saedeleer, C., Bengoetxea, A., Lipshits, M., Cebolla, A., Servais, L., Dan, B., Berthoz, A., McIntyre, J.: Effect of gravity on human spontaneous 10-Hz electroencephalographic oscillations during the arrest reaction. Brain Research 1121, 104–116 (2006)

Building Dependable EEG Classifiers for the Real World – It's Not Just about the Hardware

Gene Davis[1], Djordje Popovic[1,2], Robin R. Johnson[1], Chris Berka[1], and Mirko Mitrovic[1]

[1] Advanced Brain Monitoring, Inc. 2237 Faraday Avenue, Suite 100, Carlsbad, CA 92008
[2] University of Southern California, Los Angeles, CA 90089
{Gene,Dpopovic,Rjohnson,Chris,Mirkom}@b-alert.com

Abstract. One of the major deficiencies with the EEG-based classifiers used in today's laboratory settings is that they are often ill suited for the real world. In many cases the classifiers that were painstakingly developed in the controlled laboratory environment become unreliable with increased mobility of the user. In addition to increased mobility, many real world scenarios impose constraints on data collection that cannot be accommodated by the lab-created classifier. Addressing these issues throughout the development process of EEG-based classifiers by building hardware, software, and algorithms intended for use in the real world should result in more dependable classifiers. With this approach we were able to collect and classify data on a research vessel at sea, in the desert by night, on dismounted soldiers in the training field, and everywhere between.

Keywords: Electroencephalogram (EEG), Mobile EEG, Operational Neuroscience, Engagement, Workload, Drowsiness.

1 Introduction

Researchers have been interested in developing electroencephalogram (EEG) based classifiers for over forty years, in order to enable brain-computer interface (BCI) and neuro-feedback applications. Classifiers have been developed to track global activation state(s) as well as specific cognitive and medically diagnostic states [1-5]. Many commercially available EEG systems allow for the simultaneous recording of high quality EEG from a large number of scalp locations (128 – 256) in controlled laboratory environments and the advent of powerful digital electronics has allowed a shift in focus from simple data collection to development and implementation of complex signal processing and pattern recognition techniques that can be programmed to run in real-time. EEG metrics have been developed to quantify alertness, engagement, drowsiness and working memory, and the integration of EEG metrics into the evaluation of Human-Computer Interfaces provided the foundation for establishing new fields including neuroergonomics [1, 6-14]. However, the use of EEG outside of the lab was, until recently hampered by high susceptibility of EEG to movement artifacts, as well as environmental and physiological noise with amplitudes several orders of magnitude larger than the typical EEG amplitude (20-50 μV). Only in the last decade

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 355–364, 2009.

has EEG left the laboratory, and entered the field with the development of portable EEG systems that include the acquisition tools and noise rejection techniques necessary for high quality data collection in real world environments. Having these tools and systems enables implementation of EEG-based classifiers in real time in the real world.

Real-world, real-time EEG applications must meet certain requirements to be relevant and useful. First, the equipment, software and algorithms required for acquisition and processing must be easy to set-up and use. The hardware needs to be mobile, lightweight, robust, flexible, and reliable, whereas the acquisition and processing software should be intuitive and flexible. Finally, the algorithms should be developed with real-world applications in mind. While there are a number of potential approaches to developing real-world, real-time classifiers, at a minimum the hardware, software, and algorithm development methodology must be part of the overall game plan.

2 Methods

2.1 It's All about the Hardware

Hardware for in-field acquisition of EEG must address the following issues: portability, ease of set-up and use in the field, durability, and minimization of signal acquisition artifacts. Mobility and portability requirements translate into a need for lightweight, simplified form factors, and preferably a wireless system. Wireless systems provide maximal flexibility and platform-independence in addition to improving durability while reducing environmental artifact during acquisition. To meet ease of use requirements, it should be easy to apply by the end user and once applied, it should become "transparent" so the user is able to focus on the task at hand (not the equipment). Finally the system should be easy to trouble shoot with minimal set up time.

Advanced Brain Monitoring (ABM) has developed a system that meets these requirements. The physical system consists of 3 basic components plus electrodes: a skull cap, and electrode placement strip that is integrated into the skull cap, and an electronics headset. The skull cap and electrode placement strip components are size-matched, and sizing is derived from the Nasium-to-Occipital distance of the end user, allowing selection of small, medium or large sizes. The integrated, size-matched skull cap and electrode placement strip incorporates the occipital bone placement to ensure accurate and easy placement of the electrodes (according to the 10-20 system). This set up has the added advantage of allowing easy identification and correction of incorrect sizing that could lead to inappropriate placement of the electrodes and collection of faulty data. The number of electrodes used is also important in minimizing the set-up time, maintaining a light weight system, and reducing power consumption requirements that can decrease *portability, durability, and flexibility*. With only three electrodes, (Fz, Cz, and POz) the ABM system classifies a subject's arousal state on a spectrum from Sleep Onset to High Engagement [6, 15-25]. By adding only three additional electrodes, (F3, C3, and C4), cognitive workload states have been determined in multiple Augmented Cognition scenarios [17, 19, 26 – 30]. Finally, impedances below 5kOhms are not essential, and therefore need not delay the start of data collection. With the current ABM set-up data can be collected and classified with impedances as high as 45kOhms, allowing for much greater flexibility in real-world environments. All of the above allow the system to be set-up and begin data collection quickly in the field setting, and are particularly advantageous in time sensitive settings where access to the subject population may be limited or driven by inflexible third party events.

The system used for the majority of data collection to date is a Bluetooth (BT) wireless system that allows for true *mobility* in any setting. Wireless EEG has been acquired and passed to PDA's, desktops, laptops, and mini-laptops; allowing the user to move freely in any environment without fear of catching loose wires, while providing maximal flexibility in the set-up process. Additional mobility and comfort is found in the design of the EEG cap. Unlike traditional EEG caps, that use a chin strap to hold the EEG cap in place, the ABM system uses a skull cap design that holds the cap in place by applying tension around the head (similar to a head band) providing a greater range of normal movement. The end result is hardware that weighs less than 6 oz including the electrodes, cap, electronics and batteries in a system with maximal run times of over 10 hours on two AAA batteries.

The hardware discussed herein, has a small vertical and horizontal footprint that allows the system to be integrated into multiple configurations on the end-user. The condensed size has allowed the system to be integrated with and under, Kevlar and safety helmets, fNIR headsets, gas masks and various head mounted eye trackers [31-33]. The system design provided the flexibility that accommodated these various configurations: the flexible cap and strip material and configuration provided the robustness needed to maintain good scalp contact required for high quality data acquisition with minimal noise in the signal. The electronics that provide amplification and digitization, as well as BT transmission are encased in a plastic reinforced with Lexan material, for high impact resistance. The durability and robustness of the hardware system was perhaps best demonstrated during use in live-fire, Simunitions training exercises occurring in a rainstorm at Aberdeen proving grounds [34], where the headset sustained a direct hit, yet maintained full functionality, continuing to acquire high quality data. In the real world, a system that acquires data for a classifier will need to be ready for anything, including: rain, misuse, abuse, and neglect- and yet be able to continue to collect high quality EEG signals that are the basis of the classifier.

Ideally, the hardware design should minimally impact the state of the user; they should forget that the system is on. Depending on the level of intrusiveness of the hardware, the classifications characterizing distraction, work load, or effort described may be from the task, from coping with the data collection system, or any combination of related factors. By incorporating comfort into the original design of the ABM system the end result is a system that has been worn for 24 hours of continuous use, and has been worn by over 1000 subjects with minimal awareness of the data collection system.

2.2 It's All about the Software

Real-world software for acquisition of EEG must address the following issues: At minimum, the software required for acquisition and processing must be easy to install and intuitive to use. Ideally it should also provide mobility options, along with flexible interfaces, inputs and outputs, and reliably collect and save data. The software would be of greater assistance to the non-EEG researcher as well as the experienced EEG researcher if it were also able to identify and provide feedback regarding artifacts associated with environmental noise and/or physiological noise. Finally, the software should allow for flexible interaction with third-party software programs to ensure maximal applications of the classifiers developed.

The B-Alert Acquisition software developed by ABM was designed to meet the above needs. The software for running data acquisition in the field is less than 50 megabytes in size, can be installed in under five minutes, and does not require any additional third party software or drivers making it truly plug and play. This platform allows acquisition through a desktop, laptop, mini-computer, or PDA. As computer issues are not uncommon in real-world applications, this system allows for quick adaptation to another collection interface to be adapted as needed (either through changing to another similar system such as an additional laptop; or changing systems completely as needed such as from a laptop to a PDA). With this high degree of simplicity and flexibility, researchers are enabled to develop a plethora of experimental designs and applications.

The software has simple selection options for acquisition and retransmission allowing for observation of the signal quality from a secondary computer. In addition, prior to acquisition the software completes both an impedance check and an artifact evaluation for each electrode. Thus, the researcher can have great confidence in the initial signal quality. Finally, while in Acquisition mode(s) the software provides helpful feedback (regarding artifact as well as overall signal quality) to the end user allowing even an untrained EEG technician to quickly recognize and troubleshoot poor EEG signal quality.

The ABM B-Alert Acquisition software suite has been used to provide input from EEG into several other third party systems to enable closed-loop feedback systems. The algorithms for identifying arousal and workload states may be passed (as would any additional algorithms that may developed), as well as the raw EEG signal. These options are easily enabled, and can be used to develop and apply real-time feedback systems in field operational environments. The configurations enabled in the B-alert acquisition software allow researchers to collect the developmental data for real-world algorithms in the real world. Perhaps equally important, the system also then enables application of these algorithms (once developed and validated) in the real world. In several recent collaborations with the Lockheed-Martin Advanced Technology Laboratory, the ABM software provided outputs via DLL (digital library link) on operator levels of engagement and mental workload that were used to drive changes in the display of a Tactical Tomahawks Weapons training simulation. The resulting closed-loop system provided an 80% reduction in launch time deviations from optimal and a 55% reduction in the number of late launches [26].

2.3 Its All about the Algorithms

Real world algorithms must be flexible, robust, and developed with real-world applications in mind. To meet these requirements, the following issues will need to be considered: the design of the initial data collection experiments upon which the classifiers will be built, the feature extraction method for model building, the types of classifier models that may be considered, and the validation of the classifier. Ideally, the initial experimental design will be informed by the planned methodology for mathematically building the classifier, and these will both inform the validation process. In other words, the development of a classifier should be a well-designed rational process based on a plan that takes into account the final application(s) envisioned for the classifier.

Experimental design for initial data collection will need to address many aspects of the planned research. First, the overall design of the experiment should be considered, to ensure proper experimental control of unexplainable variance that may de-stabilize the models that are developed. Full factorial designs are not required, but well thought out protocols will ensure that the appropriate data (such as ERPs) are in fact available for classifier development. It is also essential that all subjects are exposed to identical scenarios for the purposes of development- casual changes will lead to noise that will negatively impact algorithm development. The experimental design for the initial data collection should include an acknowledgement of real-world concerns: and may benefit greatly from at least some real–world data.

Next, understanding the number and quality of anticipated features to be extracted must be taken into consideration in regard to the sample size selected. In general, stable mathematical models with good generalization ability require a minimum sample size that is at least one order of magnitude larger than the number of features extracted per unit of observation (e.g. a single-trial ERP waveform) and subsequently used for classification. Classifiers that are built on a sample of insufficient size tend to memorize rather than learn from the sample, which usually results in poor 'real world' performance on cross-validation in spite of high classification accuracy achieved during the model development. This problem is aggravated by complexity of the selected classifier (e.g. non-linear classifiers are more vulnerable than linear, neural networks with more hidden layer neurons are more easily over-trained than simpler architectures) or by non-linear transformations of the input space, typical for support vector machines (SVM), that in a non-transparent way significantly increase the real number of features (often as $\sim O(N^2)$ or $\sim O(N^3)$, where N is the number of 'visible', i.e. nominal features in the original input space). Other issues may also impact the sample size that can be reasonably collected, including access to the appropriate subjects.

Building a classifier (in a broad sense of the word) can take a number of paths, but it usually consists of three tightly connected phases: choice of feature extraction method(s), choice of feature selection method(s) and finally selection of classifier in a narrower sense of the word. Feature extraction methods serve to reduce the dimensionality of the original space containing input variables by determining an appropriate subspace of, in most cases, smaller dimensionality. Principal component analysis (PCA) is the most popular technique for dimensionality reduction, but other linear transforms (factor analysis, linear discriminant analysis, projection pursuit) or nonlinear techniques (Kernel PCA, multi-dimensional scaling, and self-organizing maps) can be used instead. One should note that for low-dimensionality input spaces feature extraction can be omitted. Feature selection methods take the original or transformed set of N input variables and down-selects a subset of M ($M<N$) variables that provide the best discrimination among the different classes into which signals (e.g. an ERP) should be classified. Exhaustive search, branch-and-bound search, sequential forward

or backward search, plus-L-take-away-R selection and sequential forward and backward floating search are some of the popular methods available to researchers. The key for a successful feature selection is the balance between the optimality of a method (i.e. how likely it is to find a local rather than the global minimum) and the speed of convergence (e.g. branch-and-bound search is optimal but very slow). Finally, a variety of techniques is available that will group similar vectors of selected features into few desired classes such as the k-nearest neighbor rule, binary decision tree, Bayes classifier, logistic classifier, Parzen classifier, Fisher linear discriminants, support vector machines and neural networks. Because of the existence of so many (in some respects overlapping) techniques and ready availability of fast computers nowadays, attempts at building a classifier easily turn into a fishing expedition. Ideally, the researcher will understand the goal of the classifier and select the appropriate techniques in order to arrive at a solid solution.

Finally, validation of any classifier must be completed prior to finalization and application. Validation may be done on the initial data set using various techniques, including bootstrapping, and hold out methods. While such an approach is statistically legitimate and can save both time and financial resources, it may not be sufficient to ensure true real-world generalized applicability. Therefore, it is highly desirable to ensure that once developed, the classifier be cross-validated on a unique dataset different from that used for the development of the classifier. Over time, occasional re-examination of the classifier's performance as it is used can also serve to ensure that it maintains real-world viability.

3 Results

Classifiers that have been successfully collected include both arousal state (sleep onset-distraction-low engagement-high engagement) and cognitive workload (low-high). Over 1000 subjects have had data collected in over 2000 sessions by over 30 in-house technicians at ABM, and our client/collaborators have collected hundreds of additional subjects with over 50 client/collaborator technicians. Through ABM work and that of our collaborators, EEG data has been collected on soldiers in the field at 29 Palms, in live-fire exercises at Aberdeen Training Grounds inNorfolk, Virginia, and Camp Pendleton (San Diego, CA), in vehicles, and in the classroom. EEG data has been collected from expert imagery analysts, expert marksman, expert chemistry

Fig. 1. From left to right: Engagement levels during marksmanship studies, data collection at 29 Palms, drowsiness classifications for 29 Palms study

students, as well as a plethora of healthy control volunteers, novice marksman, sleep disordered patients, and other experimental subjects. Some exemplar data results are shown in Figure 1.

4 Conclusion

Investigations of human mental activity have employed EEG recordings for nearly a century since the first recordings were made by Hans Berger in 1929 [35]. Today EEG is routinely used for overnight sleep studies in the laboratory and in neurology to characterize epilepsy and neurological disorders, but the great leaps in EEG research can be largely attributed to the wealth of information generated by psychologists and neuroscientists using EEG to investigate brain, mind and behavior. Although the relationships between specific mental states and EEG are just beginning to be understood, the foundation of work in detecting global state changes is sufficient to begin developing practical applications. Our team developed hardware and software to facilitate the widespread and routine use of EEG outside the laboratory supporting a growing number of applications in education and training, human factors evaluations, military operations and market research. Our team developed methods for Psychophysiological Profiling that can now be accomplished by integrating EEG and EKG with cognitive tests. The result is efficient, inexpensive assessment of alertness, attention, learning and memory, providing a quantitative profile of impairment that can be used for patients with sleep, neurological and psychiatric disorders. These NeuroAssays can be used for diagnostic and treatment outcome evaluations, pharmaceutical investigations, and to identify potential biomarkers for specific diseases [15-17].

Field applications for the EEG technology include real-time assessment of drowsiness for truck drivers or airline pilots in military, industrial or other operational settings. The integration of EEG monitoring of operator status offers the possibility of allocating tasks between machines and humans based on the operator status. Intelligent feedback or "closed-loop" systems can facilitate active intervention by the operator or through a third party (man or machine), increasing safety and productivity. Another novel approach to this evolving technology is to radically rethink the design of human-machine system interfaces to optimize the flow and exchange of data between humans and machines. The new discipline of neuroergonomics has taken on this challenge combining understanding of the neural bases of cognition and behavior with the design and implementation of technology. Our team is collaborating with educators in developing next generation EEG technology to build models of student learning using EEG for non-intrusive assessment of cognitive processes including attention, working memory, workload and problem solving [29-30]. The vision is to integrate EEG into interactive tutorials and training simulations. There are no limits to the applications of the venerable technology now that it has left the laboratory.

Although the hardware, software, and algorithms are continually undergoing changes to provide better end results, the current complete package to date has been used in a variety of locations, weather conditions, and configurations that were previously impractical if not impossible to collect EEG data in with other commercially available systems. Our system will continue to be developed to improve the durability, flexibility, and applicability in the real world, including the development of classifiers for the real world.

Acknowledgements. This work was supported by NIH\NIMH Grant No. MH078436, NIH\NHLBI Grant No. HL70484, NIH\NIMH Grant No. MH64344-01, DARPA\DSO Grant No. 20064169, DARPA Accelerated Learning Program, Dr. Amy Kruse, PM and theDARPA Augmented Cognition program, Drs. Dylan Schmorrow & Amy Kruse, PM.

"A" (Approved for Public Release, Distribution Unlimited).

References

1. Besserve, M., Jerbi, K., Laurent, F., et al.: Classification methods for ongoing EEG and MEG signals. Biological Research 40, 415–437 (2007)
2. Gevins, A.S., Morgan, N.H.: Classifier-directed signal processing in brain research. IEEE Transactions on Biomedical Engineering 33, 1054–1068 (1986)
3. Mirsky, A.F., Pragay, E.B.: The relation of EEG and performance in altered states of consciousness, vol. 45, pp. 514–534. Research Publications – Association for Research in Nervous and Mental Disease (1967)
4. Walter, D.O., Rhodes, J.M., Adey, W.R.: Discriminating among states of consciousness by EEG measurements: a study of four subjects. Electroencephalography and Clinical Neurophysiology 22, 22–29 (1967)
5. Walter, D.O., Rhodes, J.M., Kado, R.T., Adey, W.R.: A normative library of the human EEG, assessed by computer analysis in relation to behavioral states. Electroencephalography and Clinical Neurophysiology 23, 87–88 (1967)
6. Berka, C., Davis, G., Johnson, R.J., et al.: Psychophysiological Profiles of Sleep Deprivation and Stress during Marine Corps Training. Sleep 30, A132 (2007)
7. Bodenstein, G., et al.: Computerized EEG pattern classification by adaptive segmentation and probability-density-function classification. Description of the method. Computers in Biology and Medicine 15, 297–313 (1985)
8. Creutzfeldt, O.D., Bodenstein, G., Barlow, J.S.: Computerized EEG pattern classification by adaptive segmentation and probability density function classification. Clinical evaluation. Electroencephalography and Clinical Neurophysiology 60, 373–393 (1985)
9. Kramer, A., Parasuraman, R.: Neuroergonomics: Application of Neuroscience to Human Factors. In: Caccioppo, J.T., Tassinary, L.G., Berntson, G.G. (eds.) Handbook of Psychophysiology. Cambridge University Press, New York (2005)
10. Parasuraman, R.: Neuroergonomics: Research and Practice. Theoretical Issues in Ergonomics Science 4, 5–20 (2003)
11. Parasuraman, R.: Neuroergonomics: The Brain at Work. Oxford University Press, New York (2005)
12. Peltoranta, M., Pfurtscheller, G.: Neural network based classification of non-averaged event-related EEG responses. Medical and Biological Engineering and Computing 32, 189–196 (1994)
13. Pojman, N., Johnson, R.J., Kintz, N., et al.: Assessing fatigue using EEG classification metrics during neurocognitive testing. In: 23rd Annual Meeting of the Associated Professional Sleep Societies, LLC, Seattle, WA (Submitted, 2009)

14. Wilson, G., Fisher, F.: Cognitive task classification based upon topographic EEG data. Biological Psychology 40, 239–250 (1995)
15. Berka, C., Johnson, R., Whitmoyer, M., et al.: Biomarkers for Effects of Fatigue and Stress on Performance: EEG, P300 and Heart Rate Variability. Human Factors and Ergonomics Society, New York (2008)
16. Berka, C., Levendowski, D., Cvetinovic, M., et al.: Real-time Analysis of EEG Indices of Alertness, Cognition, and Memory Acquired with a Wireless EEG Headset. International Journal of Human-Computer Interaction 17, 151–170 (2004)
17. Berka, C., Levendowski, D., Davis, G., et al.: Nicotine Administration and Withdrawal Effects on EEG metrics of Attention, Memory and Workload: Implications for Cognitive Resource Allocation. In: Schmorrow, D., Stanney, K., Reeves, L. (eds.) Augmented Cognition: Past, Present and Future, Foundations of Augmented Cognition, pp. 174–183. Strategic Analysis, Inc., Arlington (2006)
18. Berka, C., Levendowski, D., Konstantinovic, Z.R., et al.: Detection of drowsiness with realtime analysis of the electroencephalogram (EEG). In: National Institutes of Health SBIR Conference, Bethesda, MD (2000)
19. Berka, C., Levendowski, D., Ramsey, C.K., et al.: Evaluation of an EEG-Workload Model in an Aegis Simulation Environment. In: Caldwell, J.A., Wesensten, N.J. (eds.) Proceedings of SPIE Defense and Security Symposium, Biomonitoring for Physiological and Cognitive Performance during Military Operations, vol. 5797, pp. 90–99. SPIE: The International Society for Optical Engineering, Orlando, FL (2005)
20. Berka, C., Levendowski, D., Westbrook, P., Davis, G., Lumicao, M., Olmstead, R., Popovic, M.V., Zivkovic, V., Ramsey, C.: EEG Quantification of Alertness: Methods for Early Identification of Individuals Most Susceptible to Sleep Deprivation. In: Caldwell, J.A., Wesensten, N.J. (eds.) Proceedings of SPIE Defense and Security Symposium, Biomonitoring for Physiological and Cognitive Performance during Military Operations, vol. 5797, pp. 78–89. SPIE: The International Society for Optical Engineering, Orlando, FL (2005b)
21. Berka, C., Levendowski, D.J., Konstantinovic, Z.R., et al.: Detection of electroencephalographic indices of drowsiness in realtime using a multi-level discriminant function analysis. Sleep 23, A243–A244 (2000)
22. Berka, C., Levendowski, D.J., Lumicao, M.N., et al.: EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. Aviation, Space and Environmental Medicine 78, B231–B244 (2007)
23. Mitler, M.M., Westbrook, P., Levendowski, D.J., et al.: Validation of automated EEG quantification of alertness: methods for early identification of individuals most susceptible to sleep deprivation. Sleep 25, A147–A148 (2002)
24. Westbrook, P., Berka, C., Levendowski, D., et al.: Quantification of Alertness, Memory and Neurophysiological Changes in Sleep Apnea Patients Following Treatment with nCPAP. Sleep 27, A223 (2004)
25. Westbrook, P., Berka, C., Levendowski, D.J., et al.: Biobehavioral quantification of alertness and memory in patients with sleep apnea. Sleep 25, A49–A50 (2002)
26. Craven, P.L., Belov, N., Tremoulet, P., et al.: Cognitive Workload Gauge Development: Comparison of Real-time Classification Methods. In: Schmorrow, D., Stanney, K., Reeves, L. (eds.) Augmented Cognition: Past, Present and Future, Foundations of Augmented Cognition, pp. 66–74. Strategic Analysis, Inc., Arlington (2006)
27. Poythress, M., Russell, C., Siegel, S., et al.: Correlation between Expected Workload and EEG Indices of Cognitive Workload and Task Engagement. In: Augmented Cognition 2006 Conference (2006)

28. Poythress, M., Russell, C., Siegel, S., et al.: Correlation between Expected Workload and EEG Indices of Cognitive Workload and Task Engagement. In: Schmorrow, D., Stanney, K., Reeves, L. (eds.) Augmented Cognition: Past, Present and Future, Foundations of Augmented Cognition, pp. 32–44. Strategic Analysis, Inc., Arlington (2006)
29. Stevens, R., Galloway, T., Berka, C.: Allocation of Time, EEG-Engagement and EEG-Workload Resources as Scientific Problem Solving Skills are Acquired in the Classroom. In: Augmented Cognition 2007 Conference (2007)
30. Stevens, R., Galloway, T., Berka, C.: EEG-Related Changes in Cognitive Workload, Engagement and Distraction as Students Acquire Problem Solving Skills. In: The User Modeling Conference, Athens, Greece (2007)
31. Dickson, B.T.: The Cognitive Cockpit - a test-bed for Augmented Cognition. In: 1st International Conference on Augmented Cognition, Las Vegas, NV (2005)
32. Kinces, W.E.: DaimlerChrysler's closed loop integrated prototype: Current status and outlook. In: 1st Augmented Cognition International, Las Vegas, NV (2005)
33. Ververs, P.M., Whitlow, S., Dorneich, M., Mathan, S.: Building Honeywell's Adaptive System for the Augmented Cognition Program. In: 1st International Conference on Augmented Cognition, Las Vegas, NV (2005)
34. Mathan, S., Whitlow, S., Dorneich, M., Ververs, P., Davis, G.: Neurophysiological Estimation of Interruptibility: Demonstrating Feasibility in a Field Context. In: Augmented Cognition 2007 Conference (2007)
35. Berger, H.: Uber das Elektroenzephalorgam des Menschen. Arch Psychiatr. Nervenk 90, 527–570 (1929)

Improved Team Performance Using EEG- and Context-Based Cognitive-State Classifications for a Vehicle Crew

Kevin R. Dixon[1,*], Konrad Hagemann[2], Justin Basilico[1], Chris Forsythe[1], Siegfried Rothe[2], Michael Schrauf[2], and Wilhelm E. Kincses[2]

[1] Sandia National Laboratories, P.O. Box 5800, Mailstop 1188, Albuquerque, NM 87185, U.S.A.
jcforsy@sandia.gov
[2] Daimler AG, Group Research, 059/G024-BB, 71059 Sindelfingen, Germany

Abstract. We present an augmented cognition (AugCog) system that utilizes two sources to assess cognitive state as a basis for actions to improve operator performance. First, continuous EEG is measured and signal processing algorithms utilized to identify patterns of activity indicative of high cognitive demand. Second, data from the automobile is used to infer the ongoing driving context. Subjects participated as eleven 2-person crews consisting of a driver/navigator and a commander/gunner. While driving a closed-loop test route, the driver received through headphones a series of communications and had to perform two secondary tasks. Certain segments of the route were designated as threat zones. The commander was alerted when entering a threat zone and their task was to detect targets mounted on the roadside and engage those targets To determine targeting success, a photo was taken with each activation of the trigger and these photos were assessed with respect to the position of the reticle relative to the target. In a secondary task, the commander was presented a series of communications through headphones. Our results show that it is possible to reliably discriminate different cognitive states on the basis of neuronal signals. Results also confirmed our hypothesis: improved performance at the crew level in the AugCog condition for a secondary communications tasks, as compared to a control condition, with no change in performance for the primary tasks.

1 Introduction

Currently, crews for military light vehicles face a significant challenge due to information overload. Within a context where one individual serves as driver and navigator and a second as commander and gunner, each crew member must continuously manage their primary and one or more secondary tasks. For instance, the driver may be required to simultaneously operate the vehicle, navigate to an objective through unfamiliar roadways and terrain, and monitor radio traffic. In addition, these tasks may

* Corresponding author.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 365–372, 2009.

need to be performed under the stress of hostile attack or improvised explosive device. For the operational effectiveness and safety of these military personnel, technology solutions are needed that allow crews to operate more effectively in multi-tasking, information-intensive, high-stress environments.

This paper describes research conducted by Daimler AG, Research Group and Sandia National Laboratories to prototype and test an augmented cognition system for enhancing team performance for crews of military light vehicles. In this system, continuous EEG is measured and signal processing algorithms utilized to identify patterns of activity indicative of high cognitive demand. Second, data from the automobile is used to infer the ongoing driving context, and corresponding levels of cognitive demand. Based upon these measures, mitigation mechanisms were initiated to lessen the cognitive demand upon crew. During the fall of 2007, experimental tests were conducted at Marine Corps facilities at Camp Pendleton, CA with U.S. Marine Corps personnel serving as experimental test participants.

2 Augmented Cognition System

For this study, a Mercedes G-Wagon was modified to serve as military-relevant experimental test platform. With respect to the exterior, this primarily consisted of an M-45 M240G 7.62 Machine gun mount on the roof of the vehicle. For experimental testing, there was no gun in the Machine gun mount.

Fig. 1. Exterior of modified Mercedes G-Wagon used in experimental testing

The interior was modified in accordance with a two-person concept of operations in which one person serves as the driver and the second the commander/gunner. For the commander's station, arm rests on either side of the passenger seat were equipped with joysticks. These joysticks provided the interface for controlling the Machine gun mount (e.g. rotating the turret, adjusting magnification, and simulated firing). A dash-mounted display provided the image from a camera positioned on the Machine gun mount which was used in aiming by placing a reticle on the desired target. In addition to the camera controls, the left joystick had two push buttons which were utilized by

Fig. 2. Interior of modified Mercedes G-Wagon to be used for experimental testing

experimental participants in performing a secondary task. Headphones were utilized to present experimental stimuli and provide sounds consistent with test scenarios (i.e. simulated gunfire).

The driver's station consisted of the standard steering wheel and steering column controls, and dash-mounted vehicle displays (i.e. speedometer). A dash mounted panel contained push buttons utilized in performing secondary tasks for the experiment. Also, as with the commander, headphones were utilized to present experimental stimuli.

3 Experimental Tasks

Subjects consisted of 22 U.S. Marine Corps active-duty personnel who participated as eleven 2-person crews. Each crew was composed of a driver/navigator and a commander/gunner, with each assigned primary and secondary tasks.

3.1 Driver Experimental Tasks

The primary task of the driver was to drive the vehicle. During the experiment, the driver operated the vehicle on a prescribed test course along existing roadways. The driver was instructed to drive no faster than 30 KPH and a governor was set to prevent the driver from exceeding this speed limit.

In addition to the primary task of driving, the driver had two secondary tasks. In one, the driver listened to reports that were presented through the headphones and categorized the reports by pressing one of five buttons mounted on the dashboard next to the steering wheel. In the other secondary task, the driver occasionally heard a call sign through the headphones and responded by pressing one of two buttons depending on whether the call sign included a designated identifier.

3.2 Commander/Gunner Experimental Tasks

To simulate the detection and engagement of adversaries, targets were placed on the roadside within so called "threat zones". The primary task of the commander/gunner was to aim and fire at targets in accordance with prescribed target sequences. The commander/gunner was also given the same call sign task as described above for the driver.

4 Augmented Cognition Mitigation Conditions

There were three experimental conditions. In the reference condition (REF), participants performed tests without the aid of mitigation. In one mitigation condition (i.e. Augmented Cognition condition, MIT), the system used the measured workload of the participants as a basis for switching tasks between crew. In a second mitigation condition (i.e. Design-based Mitigation, S_MIT) the system automatically delayed all communications that occurred when the participants were engaged in a competing task. These conditions provided two bases for comparison on the proposed system. One comparison considers mitigation versus no mitigation, whereas the second compares mitigation to a "perfect" state in which there is a priori knowledge of impending workload conditions. Each experimental condition involved one trip around the test course and took approximately 35 minutes.

To assess periods of high workload, two EEG classifiers were used:

(a) Based on 32 channel EEG input, one classifier was trained to assess the driver's workload induced by the categorization task.
(b) Based on 32 channel EEG input, the second classifier was trained to assess the commander's workload induced by the gunning task.

Difficult driving situations were determined by the context (vehicle data) classifier provided by Sandia National Laboratories.

Mitigation was applied for the call sign task. As described above, the call sign task was continuously presented to the driver and to the commander. In the unmitigated condition, a call sign was alternately presented every 30 seconds to the driver and to the commander. In the mitigated "AugCog" condition, whenever the EEG classifiers detected high workload or the context classifier detected a difficult driving situation, the call sign stream was shifted to the crew member experiencing low workload. In this case, the respective crew member received the call signs with double the usual frequency (i.e. a call sign every 15 seconds).

In the case where both crew members were in high workload, the call signs were stored in a first-in-first-out (FIFO) buffer and as soon as one of the crew members went into low workload, the FIFO buffer was emptied with the call signs stored in the buffer presented every 5 seconds.

5 Results

In assessing the performance of the EEG classifier, we considered the percentage of time in which the classifier output coincided with the experimental condition (high or low workload). Classifiers were trained for each subject. At the beginning of the experiment, the EEG data of the two crew members was recorded while they performed the same tasks as during the actual experiment. The EEG data was labeled according to the workload attributed to the tasks and fed into the classifier training algorithm. The algorithm which determines the classifier parameters with the best classification result relative to the labels was chosen. Once the classifier was determined, it was fixed and left unchanged for the remainder of the experiment.

In a second step, the classifier performance was evaluated during a reference session in which no mitigation was provided, which also served as a baseline for comparison with the mitigation session.

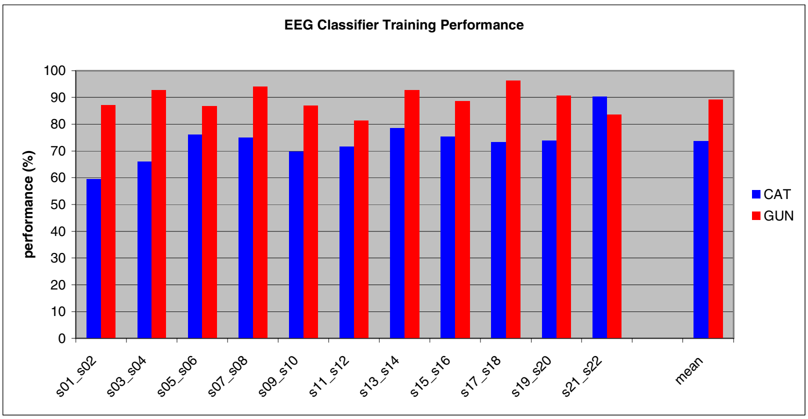


Fig. 3. EEG classifier performance (%) for all eleven teams: for the driver's categorization task (CAT, blue) and for the commander's gunning task (GUN, red) in the reference session (REF)

Detailed analysis revealed that if the EEG data from the "no mitigation" session was fed into the classifier training algorithm, usually a new classifier was obtained, leading to better classifier performances. This instability may be caused by either endogenous factors (e.g. unstable neurophysiological processes) or exogenous causes such as different experimental conditions (i.e. different rounds, different setting, variable electrode impedance, etc.).

When considering episodes in which the classifier output does not coincide with imposed experimental conditions, what may appear to be an imperfection with regard to the experimental design is always a correct decision from the classifiers point of view. It may be that the subject's workload was sometimes low during high-workload blocks or high during low-workload periods. In the current experimental setup, it was not possible to identify periods of high or low workload by means of an additional brain-based gauge that operated independent of the EEG cognitive-state classifiers.

Given the "real" cognitive state is always unknown, it is not possible to exactly determine the performance of mitigation measures. Due to an inbuilt hysteresis, as well as system intrinsic runtimes, the augmentation manager usually responded with a delay to cognitive-state changes. However, speeding up the reaction time of the system would have decreased its reliability. The configuration used in the present study was a reasonable compromise between response speed and classification reliability. Delay times after task onsets show that the system was capable of reacting within a few seconds. In cases where delays were longer, it has to be assumed that the driver actually did not experience an instantaneous cognitive state change.

5.1 Categorization Task (CAT)

Being a primary task, no changes in performance were expected between experimental conditions for the categorization task and results confirmed this hypothesis.

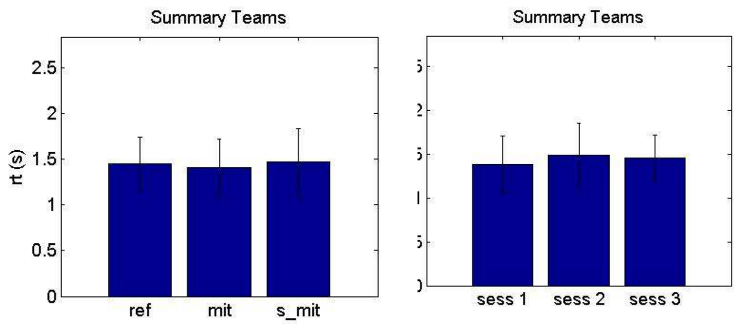


Fig. 4. Mean (n = 11) reaction times (seconds) for the categorization task (CAT) stimuli for all three experimental conditions

5.2 Gunning Task (GUN)

The gunning task (GUN) was the commander's primary task and as expected, no changes in performance were found between experimental conditions.

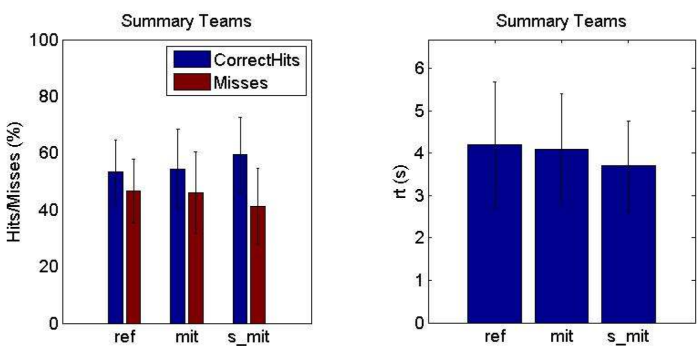


Fig. 5. Results of the GUN task plotted for three experimental conditions. (left) percent hits & misses, (right) reaction times (in seconds).

5.3 Call Sign Task (CALL)

The call sign task was the task that was mitigated and therefore represents a main outcome of the experiment. According to our hypothesis, the crew performance in this task should be better in the mitigated AugCog session (MIT) as compared to the unmitigated reference session (REF), as well as compared to the design-driven mitigation (S_MIT).

The crew performance in this task was analyzed with respect to two measures: reaction time (RT) and discriminability (d'). The results confirm the hypothesis showing that the driver's best performance was obtained in the cognitive state and driving context mitigated condition (MIT) for both reaction times (RT) and discriminability (d'). The commander's performance confirms the hypothesis in part, showing best reaction times in the MIT condition, whereas the highest d' performance was obtained in the unmitigated condition (REF). Since the call sign task (CALL) was shared by driver and commander and was mitigated between them, it is meaningful to calculate the performance measures for the entire team. This analysis shows that both performance measures are in accordance with the hypothesis, namely the reaction time is shortest and the discriminability is highest in the cognitive state and driving context based condition (MIT).

6 Conclusion

This experiment served two purposes. First, it provided an initial assessment of the viability of augmented cognition technology for military platforms. Second, the study provided experimental data to assist developers in refining the technology to enhance its potential effectiveness and relevance to military applications. This development is important because military forces are being increasingly challenged by the need to manage large volumes of information imposing high levels of cognitive load, while successfully fulfilling their assigned missions. However, it should be noted that beyond the military application targeted in these experiments, vital insights are provided for adapting the same technology for incorporation into general automotive applications providing an opportunity to enhance overall automotive safety.

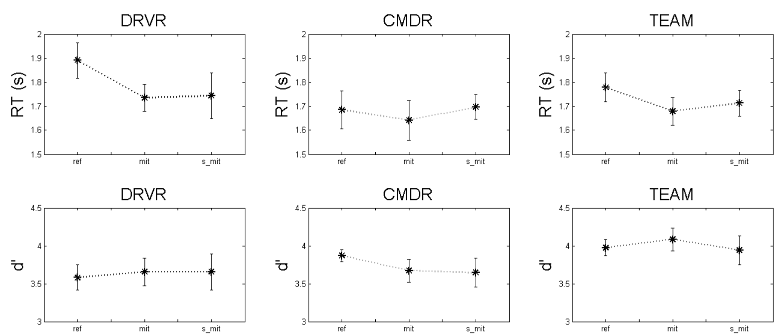


Fig. 6. Mean reaction time (RT – seconds) and mean discriminability (d' – adimensional) for driver (DRVR, left), commander (CMDR, middle) and team (TEAM, right) for three experimental conditions (REF, MIT and S_MIT)

To date, Augmented Cognition research by this team has focused on cognitive overload situations. Looking toward a complete solution for Augmented Cognition, another situation which requires understanding is the concept of task underload. Together, the measure of overload and underload, allow us a complete measure of operator "vigilance." "Vigilance" is defined as a state of mind in which a person has their attention focused on a task at a level sufficient to perform that task. When fully recognized, a vigilant system will minimize the impact of underloaded personnel while maximizing the task loading that personnel can successfully execute.

Acknowledgements

This work was supported by ONR under contract N00014-06-2-0041. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000.

Detecting Frontal EEG Activities with Forehead Electrodes

Jeng-Ren Duann[1,2], Po-Chuan Chen[2], Li-Wei Ko[2], Ruey-Song Huang[1], Tzyy-Ping Jung[1,2], and Chin-Teng Lin[2]

[1] Institute for Neural Computation, University of California San Diego, La Jolla CA, USA
[2] Brain Reseach Center, National Chiao Tung University, Hsinchu, Taiwan
duann@sccn.ucsd.edu, propc7358@gmail.com, lwko@mail.nctu.edu.tw, rshuang@sccn.ucsd.edu, jung@sccn.ucsd.edu, ctlin@mail.nctu.edu.tw

Abstract. This study demonstrates the acquisitions of EEG signals from non-hairy forehead sites and tested the feasibility of using the forehead EEG in detecting drowsiness-related brain activities. A custom-made 15-channel forehead EEG-electrode patch and 28 scalp electrodes placed according to the International 10-20 system were used to simultaneously record EEG signals from the forehead and whole-head regions, respectively. A total of five subjects were instructed to perform a night-time long-haul driving task for an hour in a virtual-reality based driving simulator comprising a real car mounted on a 6 degree-of-freedom Steward motion platform and a immersive VR environment with 360 degree projection scenes. Separate independent component analyses were applied to the forehead and whole-head EEG data for each individual subject. For the whole-head independent component (IC) set, the frontal central midline (FCM) IC with an equivalent dipole source located in the anterior cingulate cortex was selected for further analysis. For the forehead IC set, the IC with its theta power changes highly correlated with subject's driving performance was selected. The EEG power changes of the selected forehead ICs were then used to predict driving performance based on a linear regression model. The results of this study showed that it is feasible to accurately estimate quantitatively the changing level of driving performance using the EEG features obtained from the forehead non-hairy channels, and the estimation accuracy was comparable to that using the EEG features of the whole-head recordings.

Keywords: Forehead EEG, Drowsiness, Driving performance, Independent component analysis (ICA).

1 Introduction

Drowsiness or sleepiness monitoring has long been a challenging yet important application. It has been studied using video-based and/or psychophysical measurements such as eye movement, eye blink rate, heart rate, body temperature, neurophysiological signals etc. [1][2]. However, among these measures, the neurophysiological changes reflected in EEG signals seem to have early onset before the subjective

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 373–379, 2009.

symptoms of fatigue becomes manifest [3]. It has been shown that the EEG theta-band power or the combination of theta- and alpha-band powers could be highly indicative of subjects' drowsiness or sleepiness level [3][4][5][6][7][8]. Among different EEG channel locations, the frontal brain regions, especially along the frontal central midline, were reported more susceptible to the drowsiness compared to, for example, the occipital areas [1][2].

Recently, we developed and reported a prototype four-channel mobile & wireless EEG system incorporating miniature data acquisition circuitry and dry Micro-Electro-Mechanical System (MEMS) EEG electrodes with 400 ganged contacts for acquiring signals from non-hairy sites without use of gel or skin preparation. The system allows EEG monitoring of freely-moving participants performing ordinary tasks in real-world environments [9]. Because the non-hairy forehead region is usually more accessible to all types of EEG sensors and the current form factor of the abovementioned EEG system is more applicable to non-hairy sites, this study explored the efficacy of using the theta- and alpha-band spectral changes from the forehead EEG to detect drowsiness-related brain activities. The results are compared to the drowsiness-related frontal EEG spectra simultaneous recorded using a modified International 10-20 whole-head EEG system.

2 Method

2.1 Subjects

Five healthy volunteers (aged from 19 to 25 years, two females) with normal or corrected-to-normal vision participated in this study which was approved by Institutional Review Board (IRB) of National Chiao Tung University and Taipei Veterans General Hospital. All participants completed informed consent forms before being briefed on the task requirements.

2.2 Experimental Paradigm

To test the feasibility of monitoring drowsiness-related EEG activities through forehead channels, we designed and implemented a simulated hour-long nighttime long-haul highway driving task [10][11][12] on a safe yet realistic driving simulator. The driving simulator comprised a real automobile mounted on a six degree-of-freedom Stewart motion platform, installed at the center of an immersive virtual reality (VR) environment with a 360-degree display projected from seven LCD projectors. Compared to the traditional driving simulation performed in front of a computer screen, the dynamic motion platform provided multi-sensory (visual, auditory and kinesthetic) stimuli and sensation to match daily life driving experience.

The VR driving scene was designed to mimic driving at a fixed speed of 100 Km/Hr on a straight highway. Subjects were instructed to put forth their best effort to keep the vehicle cruising at the center of the fast lane. Every 5 to 10 sec, a computer-generated perturbation was randomly applied to the vehicle to simulate car randomly drifting away from the cruising position to the curb or the opposite lane (with equal probability and a constant speed of 100 km/hr) to simulate driving on non-ideal road surfaces or with poor alignment. The subjects were asked to steer the vehicle back to the previous cruising position using a steering wheel as quickly as possible. Since subject response time would be primarily affected by their drowsiness level, the longer it took for them to respond, the farther the vehicle would deviate. As a result, the deviation of vehicle (or the driving error) could be indicative of the subjects' drowsiness level.

All driving experiments were conducted in the early afternoon after lunch because during daytime sleepiness is greatest over the mid-afternoon [13]. All subjects were also asked to return for a second experiment in two weeks.

2.3 EEG Acquisition

EEG data were acquired at a sampling rate of 500 Hz using a SynAmps2 NeuroScan system (Compumendics, Ltd., VIC, Australia). Thirty (30) scalp electrodes based on a modified International 10-20 System was used to collect whole-head EEG data. However, channels Fp1 and Fp2 were left unconnected to accommodate a 3-by-5 grid of Ag/AgCl electrodes mounted on a 10 cm x 6 cm fabric patch. This setup resulted in a total of 43 EEG/EOG channels.

2.4 Data Analysis

Behavioral Data. In each hour-long EEG session, 200 ~ 300 lane-perturbation events were recorded. For each lane-perturbation event, the driving error was assessed by the maximum absolute deviation between the perturbation onset and subject's respond onset. Note that, since the vehicle might not always return to the exact center position of the cruising lane, the driving error thus could not be simply computed from the center of cruising lane. In certain circumstances, subjects might completely fell asleep and fail to respond to the deviation before the vehicle ran into the curb on the side. It was thus necessary to apply a hard limit to the driving errors. As a result, we could exclude episodes of completely falling asleep from the further analysis.

Then, the temporal profile of the driving errors was smoothed using a 90-sec square moving-average window, advancing at 2-sec step to eliminate variance with cycle lengths shorter than 1-2 minutes because the cycle of fluctuations of drowsiness level was in general longer than 4 minutes [7][8]. Finally, each driving error profile was normalized to percentile with respect to the maximum.

EEG data. The acquired EEG signals were first inspected to remove any bad EEG channels and/or bad EEG portions with unreasonable artifacts. The screened EEG signals were then band-pass filtered with cutoff frequencies of 0.5 and 50 Hz. Independent Component Analysis (ICA) was separately applied to decompose the forehead and whole-head EEG data into maximally temporally independent EEG activities. For each independent component (IC) resulted from the decomposition of the whole-head EEG data, its spatial map was subjected to source localization process implemented in EEGLAB (DIPFIT2 plug-in, http://sccn.ucsd.edu/eeglab) to find its equivalent dipole location(s) [14][15]. The IC, with its equivalent dipole located in the frontal central midline area (likely the anterior cingulate cortex, ACC), was grouped into a FCM cluster as the EEG power of this IC has been reported highly correlated with the drowsiness level [1][2]. The power spectrogram of the FCM IC time course was computed using a short-term Fourier transform (STFT) with 1-sec window and 0.75-sec window overlap. Finally, a 90-sec moving-average window was applied to the spectrogram. To verify the relationship between the power spectra of the FCM ICs and the driving performance, correlations between subjects' driving performance and the theta-band (4-7 Hz) EEG power averaged from the smoothed EEG power spectrogram were calculated.

For the EEG data obtained from the forehead electrodes, ICA resulted in 15 ICs with distinct EEG activities. Since there were no guidelines for selecting component(s) of interest for the forehead EEG, we correlated the time courses of theta-band power of each of the 15 ICs with the driving performance temporal profile and selected ICs with highest correlation between the two time courses, for each of the driving session, for further analysis.

Estimating driving performance using EEG powers. Since each subject had two experimental sessions, it was then possible to test if the EEG power changes could be used to estimate subject's driving performance. Multivariate linear regression model was applied to the theta- and alpha-band spectral time series and subject's driving performance obtained from one session and tested on a separate test session for each of the 5 subjects [4]. That is, all the coefficients of the linear regression model obtained from one session were applied to the theta- and alpha-band spectral time series of another session to estimate the drowsiness level of subject based on the following equation:

$$y = x_\alpha \beta_\alpha + x_\theta \beta_\theta + \varepsilon \tag{1}$$

where y is the driving performance in percentile and x_α and x_θ represent the average theta- and alpha-band EEG powers derived from the second session. β_α and β_θ are the regression coefficients derived from the EEG powers and driving performance data of the first session (noted as "1→2"). Correlation coefficient between the derived and true driving performance was computed to quantitatively evaluate the estimation accuracy.

3 Results

3.1 ICs of the Forehead EEG

As mentioned above, in ICA decomposition of the forehead EEG, the IC whose theta-band spectral time series most correlated with the subject driving performance was selected. Across all subjects, the selected ICs consistently maximally projected to the forehead electrode grid locations 13 and 14 (as shown in Fig. 1B). On average, the correlation coefficient between the theta spectral time series and the driving performance was 0.83 ± 0.06. This is slightly but not significantly higher than the correlation (r = 0.81 ± 0.07) between the times course of theta power of the FCM ICs obtained from the decomposition of the whole-head EEG and the driving performance. Fig. 1A shows the mean component map or topography of the FCM ICs.

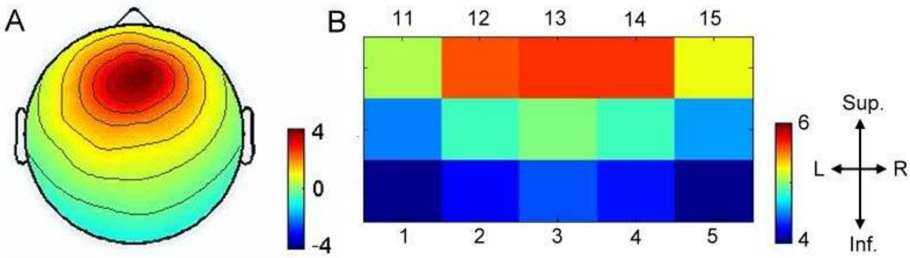


Fig. 1. A. The IC map (topography) of the frontal central midline (FCM) cluster obtained by averaging all the ICs in the FCM cluster from multiple subjects. **B.** The "topography" of the IC decomposed from the forehead EEG, which is maximally correlated with the driving performance. This is the IC we then used in predicting the driving performance based on EEG activities. As can be seen, this IC maximally projected to the electrode grid channels 13 and 14, which are located at the center of the top row on the forehead EEG channels.

3.2 Estimating Driving Performance Using EEG Powers

Table 1 summarizes the accuracy of estimating driving performance using a linear combination of theta- and alpha-band powers of the ICs obtained from either the forehead ("FH") or the whole-head ("FCM" IC) EEG. The accuracy was assessed by the correlations between the estimated and the true driving performance. The columns labeled "1→2" represents the estimating results using the linear regression model trained on Session #1 and tested on a separate test session (Session #2) for each of the 5 subjects. On average, the goodness-of-fit was 0.89 ± 0.05 for the drowsiness estimation using the combined EEG spectra from the forehead ICs, and 0.84 ± 0.07 using the EEG features from the FCM cluster derived from the whole-head EEG. The difference in the estimation accuracy, however, was not statistically significant (p = 0.08, paired t-test).

Table 1. Predicting subjects' driving performance using theta- and alpha-band EEG powers derived from the forehead ICs and FCM ICs obtained from the decomposition of the whole-head EEG

	Subj. 1		**Subj. 2**		**Subj. 3**		**Subj. 4**		**Subj. 5**	
	1→2	**2→1**	**1→2**	**2→1**	**1→2**	**2→1**	**1→2**	**2→1**	**1→2**	**2→1**
FH	0.87	0.94	0.89	0.78	0.90	0.91	0.91	0.86	0.94	0.91
FCM	0.82	0.91	0.91	0.85	0.89	0.92	0.76	0.78	0.86	0.73

The correlation between the component activations of the drowsiness-related ICs of the forehead EEG and those of the IFM ICs were also calculated (r = 0.64 ± 0.12). Since the component activities of FCM ICs mainly originated from the ACC, the high correlation between the forehead and frontal ICs suggested the selected forehead IC might mainly account for the activities from the region.

4 Discussion

In general, the combination of theta- and alpha-band powers of the ICs obtained from either the forehead or the whole-head EEG could be used to accurately estimate the driving performance (putative subject drowsiness level). The driving errors simply reflected the late responses to the vehicle deviation induced by random drifts applied to the moving vehicle. Although such driving-error index might not be a direct measure of drivers' drowsiness level, it has been proven a close approximation in the previous studies [6][16][17][18].

Cajochen and colleagues demonstrated that the low frequency (1-7 Hz) of frontal EEG signals highly correlated with the increase of sleepiness due to sustained wakefulness [1] or the administration of melatonin [3]. Such a relation has been drawn based on the associations between the frontal EEG activities and slow eye movements, eye blink rate, as well as circadian rhythm of plasma melatonin measured from the subjects. These physiological measures could be considered direct links to the drowsiness or sleepiness as defined in [1]. The current study also found a strong correlation between the frontal EEG activities and driving performance, indirectly validating the use of the driving performance as an index of subject drowsiness level.

The spectra of the ICs obtained from the decomposition of the forehead EEG could be used to estimate the driving errors as well as, if not better than, using the spectra of FCM ICs (cf. Table 1). Furthermore, the component maps of the selected forehead ICs consistently maximally projected to only one or two forehead channels (#13 or 14 in Fig. 1), suggesting drowsiness-related EEG activities might be obtained from as few as two non-hair forehead electrodes. However, higher-density forehead electrodes might be useful for ICA in separating brain activities from noises or artifacts arising from eye blink, eye movements, artifacts, muscle movements, etc.

The high correlation between the time courses of most drowsiness-relevant forehead IC and the time courses of FCM IC suggest that the EEG electrodes placed on the non-hairy forehead region might be sufficient to assess informative brain activities from the anterior cingulate cortex. In addition, this might also imply it is possible to use the forehead EEG channels to detect the brain processes in the prefrontal cortex or even other frontal regions, such as the dorsal/ventral lateral prefrontal cortices or medial prefrontal regions. As a result, the forehead EEG can be might be informative and useful in assessing the brain activities associated with many different cognitive functions, such as attention related processes, central executive functions, etc.

As the non-hairy forehead region is usually easily assessable by dry MEMS or any other types of electrodes, compared to the scalp locations covered by the hairs, results of this study might lead to broader applications in human-machine interface/ interaction design.

References

1. Cajochen, C., Khalsa, S.B., Wyatt, J.K., Czeisler, C.A., Dijk, D.J.: EEG and ocular correlates of circadian melatonin phase and human performance decrements during sleep loss. Am. J. Physiol. 277(3 Pt 2), R640–R649 (1999)

2. Cajochen, C., Knoblauch, V., Kräuchi, K., Renz, C., Wirz-Justice, A.: Dynamics of frontal EEG activity, sleepiness and body temperature under high and low sleep pressure. Neuroreport 12(10), 2277–2281 (2001)
3. Cajochen, C., Kräuchi, K., von Arx, M.A., Möri, D., Graw, P., Wirz-Justice, A.: Daytime melatonin administration enhances sleepiness and theta/alpha activity in the waking EEG. Neurosci. Lett. 207(3), 209–213 (1996)
4. Eoh, H.J., Chung, M.K., Kim, S.: Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. International Journal of Industrial Ergonomics 35(4), 307–320 (2005)
5. Horne, J.A., Reyner, L.A., Barrett, P.R.: Driving impairment due to sleepiness is exacerbated by low alcohol intake. Occup. Environ. Med. 60(9), 689–692 (2003)
6. Horne, J.A., Baulk, S.D.: Awareness of sleepiness when driving. Psychophysiology 41(1), 161–165 (2004)
7. Jung, T.P., Makeig, S., Stensmo, M., Sejnowski, T.J.: Estimating alertness from the EEG power spectrum. IEEE Trans Biomed. Eng. 44(1), 60–69 (1997)
8. Makeig, S., Jung, T.P.: Changes in alertness are a principal component of variance in the EEG spectrum. Neuroreport 7(1), 213–216 (1995)
9. Lin, C.T., Ko, L.W., Chiou, J.C., Duann, J.R., Huang, R.S., Liang, S.F., et al.: Noninvasive Neural Prostheses Using Mobile and Wireless EEG. Proceedings of the IEEE 96(7), 1167–1183 (2008)
10. Huang, R.S., Jung, T.P., Duann, J.R., Makeig, S., Sereno, M.I.: Imaging brain dynamics during continuous driving using independent component analysis. In: 35th Annual Meeting of Society for Neuroscience, Washington D.C (2005)
11. Huang, R.S., Jung, T.P., Makeig, S.: Multi-scale EEG brain dynamics during sustained attention tasks. In: Proc. IEEE ICASSP 2007, pp. 1173–1176 (2007)
12. Huang, R.S., Jung, T.P., Makeig, S.: Event-related brain dynamics in continuous sustained-attention tasks. Augmented Cognition, 65–74 (2007)
13. Roth, T., Roehrs, T., Caeskadon, M.A., Dement, W.C.: Daytime sleepiness and alertness. In: Kryger, M., Roth, T., Dement, W.C. (eds.) Principles and Practice of Sleep Medicine, 2nd edn., Philadelphia, W.B. Saunders, pp. 40–49 (1994)
14. Delorme, A., Makeig, S.: EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134(1), 9–21 (2004)
15. Oostenveld, R., Praamstra, P.: The five percent electrode system for high-resolution EEG and ERP measurements. Clin. Neurophysiol. 112(4), 713–719 (2001)
16. Lin, C.T., Chen, Y.C., Huang, T.Y., Chiu, T.W., Ko, L.W., Liang, S.F., et al.: Development of wireless brain computer interface with embedded multitask scheduling and its application on real-time driver's drowsiness detection and warning. IEEE Trans. Biomed. Eng. 55(5), 1582–1591 (2008)
17. Lin, C.T., Wu, R.C., Liang, S.F., Chao, W., Chen, Y.C., Jung, T.P.: EEG-based drowsiness estimation for safety driving using independent component analysis. IEEE Transactions on Circuits and Systems I: Regular Papers 52(12), 2726–2738 (2005)
18. Lin, C.T., Chung, I.F., Ko, L.W., Chen, Y.C., Liang, S.F., Duann, J.R.: EEG-based assessment of driver cognitive responses in a dynamic virtual-reality driving environment. IEEE Trans. Biomed. Eng. 54(7), 1349–1352 (2007)

The Effectiveness of Feedback Control in a HCI System Using Biological Features of Human Beings

Mariko Funada[1], Miki Shibukawa[2], Yoshihide Igarashi[3], Takashi Shimizu[1], Tadashi Funada[4], and Satoki P. Ninomija[5]

[1] Hakuoh University, Department of Business Administration, Oyama, Tochigi, Japan
[2] Hakuoh University, Department of Education, Oyama, Tochigi, Japan
{funada,msowlets}@hakuoh.ac.jp
[3] Gunam University, Professor Emeritus, Kiryu, Gunma, Japan
fiftystorms@lilac.plala.or.jp
[4] Rikkyo University, College of Science, Toshima-ku, Tokyo, Japan
[5] Aoyama Gakuin University, Professor Emeritus, Sagamihara, Kanagawa, Japan

Abstract. The purpose of this paper is to clarify the brain activities of human beings engaged in their tasks. Response time, correctness ratios, and Event Related Potentials (ERPs) are useful indexes of the brain activities of a subject at his task in the experiments. We analyze these indexes by a method called the Principle Component Analysis. Then we characterize the brain activities while he is engaged in the task. Finally we discuss the effectiveness of feedback control in a HCI system using these indexes.

Keywords: event related potentials, feedback control, principle component analysis, and response time.

1 Introduction

The purpose of this paper is to clarify the relationship among various indexes obtained from brain activities. We examine whether the indexes may be useful to improve the effectiveness of the feedback control in a HCI (Human Computer Interaction) system. The system consists of a subject (a human being), a computer, a display and a keyboard, where Event Related Potentials (ERPs for short) [1,2,4,6-8] are used as important information extracted from the brain. ERPs are taken from electroencephalograms (EEGs for short) [5] of the subjects.

A subject is involved in a series of tasks such that he is asked to choose the correct one from three choices shown in the display. ERPs of the subject are closely related to his brain activities when he is engaged in the multi-choice tasks. The change of ERPs during the execution of the tasks may reflect the change of physiological and/or psychological conditions of the subject as well as the laboratory situation.

We consider that the display of tasks and the ERPs of a subject are an output from the HCI system and feedback signals to the HCI system, respectively. The computer in the system may try to adjust the size of characters in the display, time duration, the format of the display and others so that the subject can be comfortably engaged in the

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 380–389, 2009.

experiments. The HCI system is depicted in Fig. 1, where two types of feedback lines, feedback (*a*) and feedback (*b*) are mainly used.

We intend to show that the HCI system can be adaptive with the conditions of the subject as well as the situation of the laboratory. We had a number of experiments to evaluate the effectiveness of the feedback control of the HCI system. This research may be applied for the design of HCI systems coping with other types of biological information.

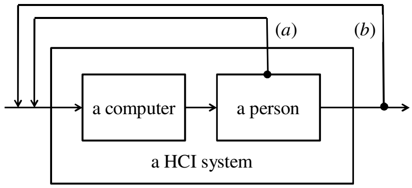


Fig. 1. An image of a feedback control HCI system

2 Experiments and Data Analysis

We repeat the following experiments from four to ten times:

1) The subjects: One is 20 years old, and the other is 21 years old. Both are male. We use "*sub* 1" and "*sub* 2" to identify them.
2) The place of the experiments: The laboratory of the first author at Hakuoh University.
3) Stimuli: We use 47 kinds of stimuli. A stimulus shown in the display is a contour of the geographical shape of a prefecture together with three choices about district capitals or prefecture capitals (see Fig. 2). We prepare three different sized stimuli for each case, a large sized one (320 ×275 pixels), a medium sized one (260×220 pixels), and a small sized one (200×160 pixels).

Fig. 2. An example of stimuli; the contour shows the shape of Tokyo, and three words of Chinese characters are *Shinjuku*, *Osaka* and *Nagano*

4) Tasks: Three types of tasks are used. These are denoted by *Task A* and *Task B*, and *Task C*. For *Task A*, a subject watches the middle sized geographical shape of a prefecture shown in the display. Then he chooses its district capital or prefecture capital from the three choices. For *Task B* and *Task C*, the actions of a subject are almost the same as the actions for *Task A*.

5) Display of stimuli: For *Task A*, a sequence of middle sized stimuli is displayed in a CRT (Cathode Ray Tube) of 19 inches placed in front of a subject. A sequence of 47 stimuli is called a set of stimuli. For *Task A*, five sets of stimuli are executed in one day. For *Task B*, three sets of stimuli are executed in one day. The small sized stimuli, the middle sized stimuli, and the third sized stimuli are used in the first set, in the second set, and in the third set, respectively. For *Type C*, the size of each stimulus in a set is randomly chosen. Such a set of stimuli is repeated three times. For any sized stimulus displayed in the CRT, the subject can watch it without moving his eyes.
6) Time duration for the stimulus display: Each stimulus, such as in Fig. 2, is displayed for 1 second. The interval between two consecutive stimuli is randomly chosen within the range 400 ms to 600 ms.
7) Time duration of an experiment: About 2 minutes are spent for a set of stimuli. The subject takes a minute interval between two consecutive sets of stimuli. Consequently, for *Task A* the time duration of 5 sets of stimuli excluding the interval time is about 10 minutes. The time duration excluding the interval time for *Task B* or *Task C* is about 6 minutes.
8) EEGs: Single polar eight channels of “International 10-20 methods” are used for the measurement of EEGs. The positions of the measurement are at Fp_1, Fp_2, C_3, C_4, O_3, O_4, C_z, and P_z. The base is A_1 that is connected to A_2.
9) The sampling frequency for A/D: 1 kHz.

We process the recorded EEGs to obtain ERPs in the following way:

1) The recorded EEGs are filtered by an adaptive filter [3] designed and made by the first author.
2) The filtered data are normalized by the average and the standard deviation of the data.
3) The normalized 47 EEGs are averaged to obtain an ERP evoked by experiments of *Task A*, *Task B*, or *Task C*. We use $ERP_{kj}(t)$ to indicate the obtained ERP, where k denotes the type of tasks, j denotes the order in the repetition of stimuli sets, and t is time [ms] (k=A, B, C, j=1,2,…,5, t=1,2,…,1000).
4) Since we repeated an experiment from four to ten times, we take the average of the 4 to 10 sampling data for $ERP_{kj}(t)$ to obtain a typical value for an ERP. We use $SERP_k(t)$ to indicate the average.

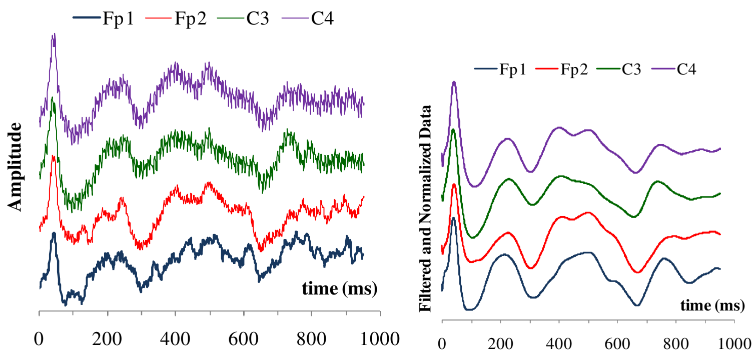
We investigate the relationship among the *response time*, the *frequency of correct answers* (i.e., the ratio of correct answers), the *latency* of ERPs, and the *amplitudes* of ERPs by analyzing the data obtained in the experiments.

3 Results

3.1 Recorded Data and Filtered Data

Examples of EEGs that are measured from *sub* 1 are shown in Fig. 3 (i). The time elapse [ms], since a stimulus is given, is shown on the horizontal axis. The amplitude of measured data is plotted in the vertical direction. In Fig. 3 (i), the lowest, the second lowest, the third lowest and the highest waveforms are plotted data from Fp_1, Fp_2,

C_3 and C_4, respectively. The measured data contain 50 [Hz] and other types of noises. These noises would be caused by electromyography, blinks or body movements of the subject and others. Before starting the experiments, the subject is asked that he should make an effort to minimize his blinks and body movements. Noises of higher frequency than frequency of EEGs are also minimized. The recorded data are filtered and normalized. These data are given in Fig. 3 (ii).



(i) Recorded EEGs (ii) Filtered and normalized data

Fig. 3. The recorded data, and the filtered and normalized data obtained from *sub* 1 during the second task (*Task A*) of the first set in the first experiment

3.2 Typical ERPs for Tasks

The results obtained from *sub* 1 are mainly described hereafter. He took the one-day experiment of *Task A* 10 times, the one-day experiment of *Task B* 4 times, and the one-day experiment of *Task C* 6 times. Consequently he took 47×5×10 stimuli of *Task A*, 47×3×4 stimuli of *Task B*, and 47×3×6 stimuli of *Task C*. First he took a one-day experiment of *Task A* once a week. Three months later after the end of the experiments of *Task A*, he took a one-day experiment of *Task B* and a one-day experiment of *Task C* in the same day once a week.

The waveforms shown in Fig. 4 are examples of $SERP_k(t)$ (k = *A, B*, *C* and t=1,2,…,1000) obtained by averaging all filtered and normalized data. These waveforms can be considered typical ERPs for *Task A*, *Task B* and *Task C*. See the potentials, P_{100}, N_{200}, P_{300} and N_{400} of the waveforms in Fig.4, where P means a positive potential and N means a negative potential. The suffix of each of these symbols indicates its latency from the start of the stimulus. Potentials P_{100}, N_{200} and P_{300} appear clearly in the waveforms for any of *Task A*, *Task B* and *Task C*. On the other hand, potential N_{400} appears clearly in the waveforms for *Task B* and *Task C* but not for *Task A*.

3.3 Comparison among Tasks

In Fig. 5, we show the tendency of how the "*average of Response time*" and "*frequency of correct answers"* for *Task A* change by repeating the experiments. Each

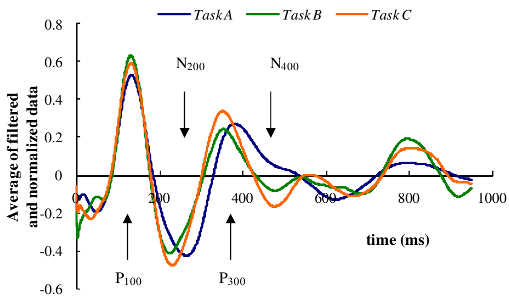


Fig. 4. Typical ERPs ($SERP_k(t)$, k=*A*, *B*, and *C*) for *Task A*, *Task B*, and *Task C*

point in the graph indicates a result for a set of tasks. Five consecutive points linked by lines are a result for a one-day experiment. The result for the 6th day is not shown in the graph, because we lost the experimental data for the 6th day. The "*Frequency of correct answers*" approaches 100% by repeating the experiments. It becomes about 97% during the last four days. The "*average of Response time*" is shortening by the repetition, but its tendency is not clear compared with the "*frequency of correct answers*".

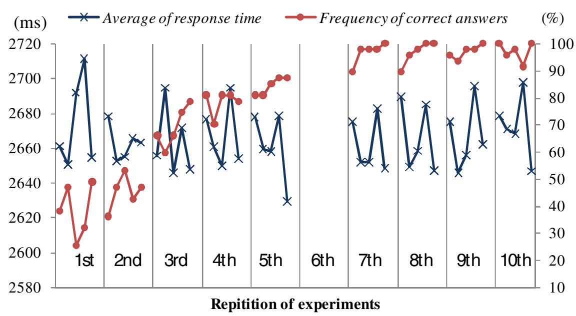


Fig. 5. Tendency of how the "*average of response time"* and the "*frequency of correct answers*" change

Let us examine how ERPs are affected by the repetition of the experiments and by the task types (*Task A*, *Task B* and *Task C*). When we use the *averaged method*, as a model of the relation between EEGs and ERPs, the following equation is widely used:

$$EEG(t) = ERP(t) + Ns(t) \quad t=1, 2, \ldots, 1000 \tag{1}$$

In equation (1) above, $Ns(t)$ is random noise including potentials caused by something unrelated to the tasks in the experiments. The average of $Ns(t)$ is 0, but it is not pure white nose. It contains EEGs caused by unexpected matters in the laboratory, the condition of the subject and others. Let us consider the EEG distribution within the whole time range. We count the number of data satisfying the following inequality (2), where the normalized values for EEGs are used:

$$\text{We count the number of positive data such that } EEG(t) > 0.5 \text{ for } 1 \leqq t \leqq 1000 \tag{2}$$

Since we first normalize the filtered data, the average and the variance of the data distribution of EEGs are 0 and 1, respectively. If the amplitude date of EEGs show *normal distribution* $N(0, 1)$, the probability $p(x > 0.5)$ is nearly 0.3. We count the number of data that satisfy inequality (2). These numbers for the 1st day experiment and the 5th day experiment are plotted as vertical values in the graph of Fig.5. In one day experiment, 235 stimuli (47 stimuli×5 sets = 235 stimuli) are given in the display. Each stimulus lasts 1000 ms. The horizontal axis of the graph represents the elapse from the start of the stimulus (0 to 1000 ms). The arrows in the graph indicate the peak values.

Let *pt* be the time point nearest to 300 ms such that number of the EEG data is a peak at *pt* (see the time points indicated by the arrows in Fig. 6). Notation $EEG_i(t)$ means $EEG(t)$ for the *i*-th stimulus among the 235 stimuli. We categorize $EEG_i(t)$ ($i = 1, 2, \ldots, 235$) into 3 classes according to the following rule (3):

$$\begin{array}{l} \text{If } EEG_i(pt) > 0.5, \text{ then } EEG_i(t) \text{ belongs to } \textit{class} \text{ I} \\ \text{else if } EEG_i(pt) >= 0.0 \text{ then } EEG_i(t) \text{ belongs to } \textit{class} \text{ II} \\ \text{else } EEG_i(t) \text{ belongs to } \textit{class} \text{ III} \end{array} \tag{3}$$

For the 5th day experiment, the numbers of elements in *class* I, in *class* II and in *class* III are 144, 33 and 58, respectively. For each class, the averaged $EEG_i(pt)$ of the 5th day experiment is shown in Fig. 7. For each t (1□t□100), $ERP(t)$ is the weighted average of these data of $EEG_i(t)$. We first calculate the average of EEGs in each class. Then $ERP(t)$ is calculated as the weighted average of the averages of these three classes. As shown in Fig. 7, the data in *class* I has a clear peak P_{300}, but P_{300} for the data in *class* II is much lower than P_{300} for the data in *class* I. The latency of P_{300} for the data in *class* III is longer than the latency of P_{300} for other classes. In Fig. 7, we notice that there is a small positive peak after P_{300} for the data in *class* III.

We choose all $EEG(t)$'s in *class* I, and then calculate the average of these data for each type of tasks (*Task A*, *Task B*, *Task C*). The average of $EEG(t)$'s calculated in

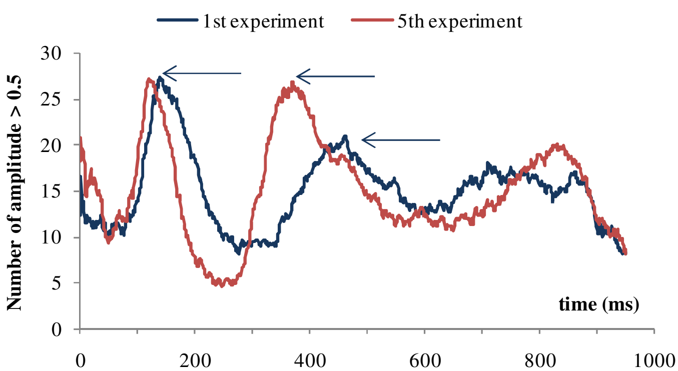


Fig. 6. The distribution of normalized EEGs greater than 0.5

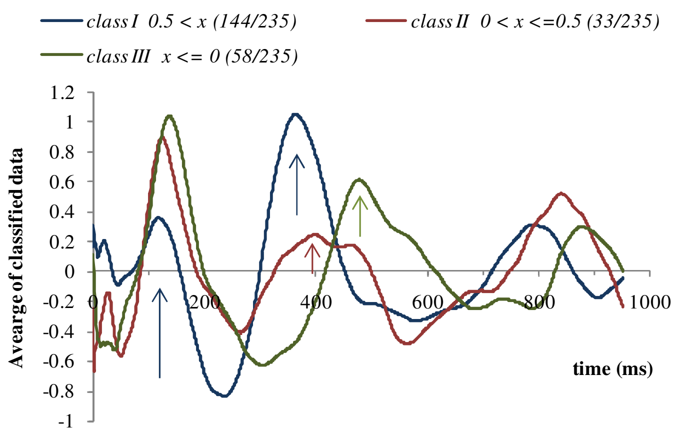


Fig. 7. The average of $EEG_i(t)$ for each of *class* I, *class* II and *class* III

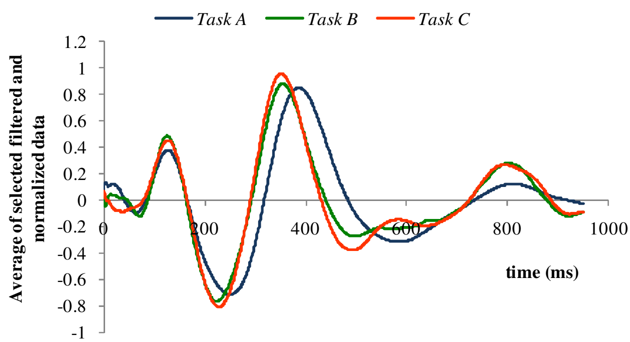
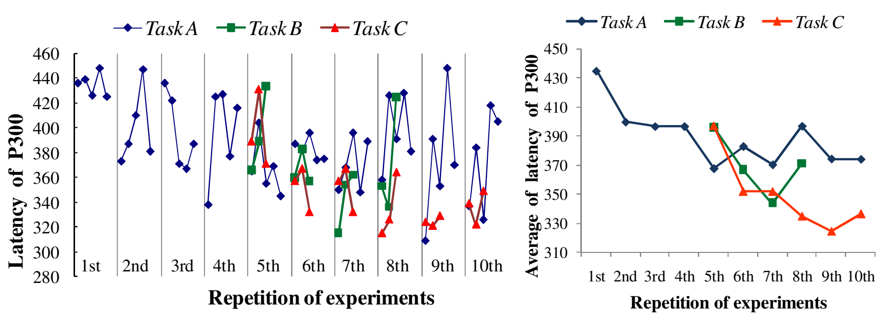


Fig. 8. Typical *SERP*(*t*)'s calculated from *EEG*(*t*)'s in *class* I



i) Change of the latency of P_{300} ii) Average latency of P_{300}

Fig. 9. The change of the latency of P_{300} through all the experiments. The horizontal axis is the order of the experimental days, and the vertical axis is the latency of P_{300}.

this way is denoted by *SERP*(*t*). An example of *SERP*(*t*) for each type of tasks is shown in Fig. 8. A peak potential P_{300} is considered to be caused by brain activities for recognition and judgment. The feature of the *SERP*(*t*) is the clear appearance of P_{300}. Through all the experiments, the change of the latency of P_{300} for each type of tasks is shown in Fig. 9 (i). The average latency for each day and for each type of tasks is shown in Fig. 9 (ii). We carried out the experiments for *Task A* before the experiments for *Task B* and *Task C*. For the first three or four days, the latency of P_{300} for *Task A* is longer than the corresponding latency for every type of the experiments after 4th day. From this tendency, we consider that the subject needs a few days to improve his skill for recognition and judgment in the experiments. The shorter latency of P_{300} reflects the better ratio of correct answers.

4 Considerations

4.1 Categories of the Human System Status

The relation between the latency and the amplitude of P_{300} is shown in Fig. 10 and Fig. 11. The horizontal axis and vertical axis in Fig. 10 are the amplitude of the average of normalized data and the averaged latency of P_{300}, respectively. Each of small circle points, small square points and small rectangle points show the result for a task of *Task A*, *Task B* and *Task C*, respectively. In Fig. 10, S_1 contains the result for *Task A* obtained in the first experiment, circle S_2 contains the results for *Task A* in the next three-days experiments, and so on. The arrows connecting these circles show the order of the experiments. Circle S_1' contains the results obtained in the first experiment for *Task B* and *Task C* (i.e., in the 5th day experiment). Circle S_2' contains other results. The results in S_2 *and* S_2*'* are resemble each other. *The results in* S_3 and S_2*'* are resemble each other. However, for the latency, the results in S_2, and S_4 are not much different. We can consider that the path indicated by the arrows in Fig. 10 shows the process of the change of the brain status. From such consideration of the categories, the learning process of the subject for *Task B* and *Task C* seems to be different from the learning process for *Task A*.

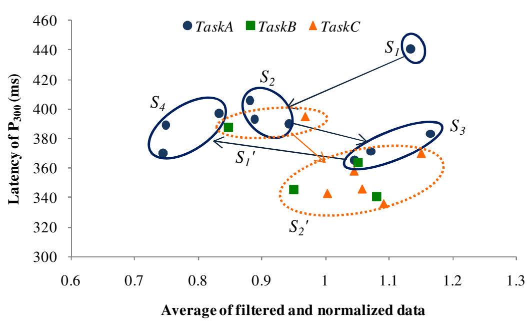


Fig. 10. The change of the average and the latency of P_{300} during the experiments

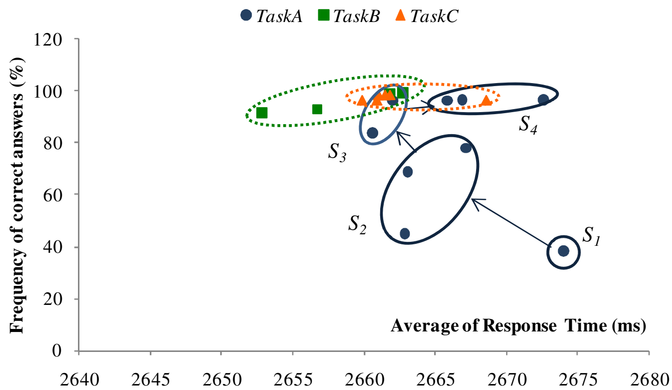


Fig. 11. The relation between the average *response time* and the *frequency of correct answers* for each of the task types (*Task A*, *Task B* and *Task C*)

The relation between the average *response time* and the *frequency of correct answers* for each of *Task A*, *Task B* and *Task C* in Fig. 11. The path indicated by the arrows in Fig. 11 shows the tendency of how the relation changes through the experiments. As shown in Fig. 11, we cannot simply say that the *response time* for *Task A* monotonically decreases by iterative learning. If we only use the data such that the *frequency of correct answers* is less than 90%, the correlation between the *response time* and the *frequency of correct answers* is about – 0.596. On the other hand, for the data such that the *frequency of correct answers* is greater than 90%, the correlation between them is about 0.502.

By applying the *Principal Component Analysis* to the values for four indexes (response time, frequency of correct answers, latency, amplitude of ERPs), we obtain the results given in Fig. 12.

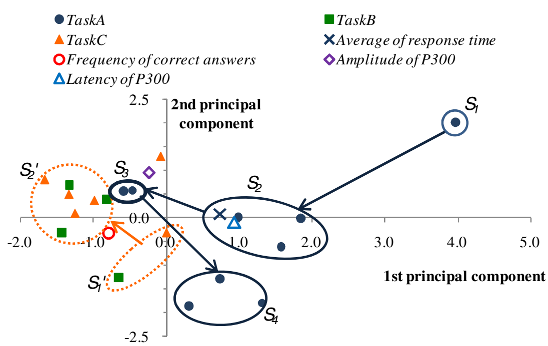


Fig. 12. The Results obtained by the *Principal Component Analysis*

5 Concluding Remarks

A HCI system can be considered a closed controlled system consisting of human beings and a computer system. The computer system is relatively stable compared with human beings. The brain activities are sensitive to the change of various factors. To realize a comfortable HCI or to improve the stability of a HCI system, stimuli and instructions given to human beings from the compute system should be well adapted to the brain status of the human beings. From our experiments and data analysis given in this paper, we can say that ERPs would be useful information to adjust stimuli and/or instructions from a computer system to human beings. ERPs calculated from EEGs of human beings could be used as feedback signals in a HCI system.

References

1. Davis, H., Davis, P.A., Loomis, L., et al.: Electrical reaction of the brain to auditory stimulation during sleep. J. Neurophysiol. 2, 500–514 (1939)
2. Dawson, G.: Cerebral responses to electrical stimulation of peripheral nerve in man. J. Neurol Neurosurg Psychiat. 10, 137–140 (1947)
3. Funada, M.F.: An Object-oriented Filtering Method for Measuring Event. The Hakuoh Graduate School Business Review 5, 93–109 (2005)
4. Luck, S.J.: An Introduction Event-Related Potential Technique. The MIT Press, Cambridge (2005)
5. Ookuma, T.: Clinical Electroencephalography, Igakusyoin (1999) (in Japanese)
6. Picton, T.W., et al.: Guideline for using human event-related potentials to study cognition: recording standards and publication criteria. Psychophysiology 37, 128–152 (2000)
7. Tanba, S., Turu, N.: Event Related Potential, Sinkouigaku-syppansya (1997) (in Japanese)
8. Turu, N.: Electroencephalography and Analysis of Electroencephalogram, Sinkouigaku-syppansya (2000) (in Japanese)

Bayesian Reconstruction of Perceptual Experiences from Human Brain Activity

Jack Gallant[1,2], Thomas Naselaris[1], Ryan Prenger[3], Kendrick Kay[2], Dustin Stansbury[4], Michael Oliver[4], An Vu[5], and Shinji Nishimoto[1]

[1] Program in Neuroscience,
[2] Departments of Psychology
[3] Physics
[4] Vision Science and
[5] Bioengineering,
University of California at Berkeley
Berkeley, CA 94720, USA
gallant@berkeley.edu

Abstract. A method for decoding the subjective contents of perceptual systems in the human brain would have broad practical utility for communication and as a brain-machine interface. Previous approaches to this problem in vision have used linear classifiers to solve specific problems, but these approaches were not general enough to solve complex problems such as reconstructing subjective perceptual states. We have developed a new approach to these problems based on quantitative encoding models that explicitly describe how visual stimuli are (nonlinearly) transformed into brain activity. We then invert these encoding models in order to decode activity evoked by novel images or movies, providing reconstructions with unprecedented fidelity. Here we briefly review these results and the potential uses of perceptual decoding devices.

Keywords: Bayesian, vision, brain-machine interface, brain-computer interface, brain reading.

1 Introduction

The human brain consists of dozens of anatomically and/or functionally distinct processing systems: sensory systems (vision, audition, etc), cognitive systems, motor systems and so on. Each subsystem is itself composed of several to several dozen distinct areas. For example, the visual system consists of at least 30 distinct areas, and each of these areas plays a unique functional role in visual perception [9]. Most areas within this complex network can be affected by both external input and output (the sensory and motor worlds) and by internal mental states (intention, attention, memory and beliefs).

The architecture of the brain suggests that each sensory or cognitive experience will be reflected in a unique pattern of brain activity that will reflect the combined influence of the external world and internal mental states. To the degree that these

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 390–393, 2009.

patterns of activity are deterministic and systematic, it should therefore be possible in principle to relate measured brain activity to sensory stimuli, motor actions and internal mental states. By implication, given a measured pattern of brain activity it should be possible to decode the corresponding subjective perceptual experiences and internal mental states.

The goal of a perceptual brain reading system is to use measurements of brain activity to decode the immediate contents of visual perception, or to decode subjective perceptual states such as visual imagery and dreaming [7]. A device of this sort could be used to solve pressing problems in many fields that require good measurements of sensory processing, including psychiatry, neurology, psychology, counseling, marketing, entertainment and the legal system.

Sensory brain decoding has traditionally been viewed as a categorization problem [4]: given a stimulus drawn from a small number of known categories (e.g., faces, places, animals), use measured brain activity to determine the specific category from which the image was drawn. This is usually accomplished by constructing a linear classifier that can discriminate the brain activity evoked by each stimulus class. Several previous studies have demonstrated that perceptual classification is possible when the classifier was trained previously on the specific classes that are to be identified [1, 2, 3, 5, 6]. However, the standard classifier approach cannot be used to classify images that belong to novel categories of stimuli that were not used to train the classifier. This is because conventional linear classifiers are not based on an explicit model that describes how the visual system encodes sensory information. Instead, they simply operate on labels that are assigned to the images. For the same reason, a classifier only enables classification of perceptual states, but does not permit reconstruction of the stimulus that evoked measured brain activity.

2 Bayesian Decoding Based on Explicit Encoding Models

We have pioneered an alternative approach to brain reading in which the decoding algorithm is inferred from one or more explicit encoding models [7]. These encoding models describe the systematic relationship between visual stimuli and brain activity. They are based on our understanding of visual function, gained from neurological, electrophysiological and neurophysiological experiments. To optimize the quality of decoding, encoding models are fit to each observer individually. In a typical experiment, several hours of functional MRI (fMRI) data are collected from a single observer while s/he watches flashed natural images or natural movies. These data are used to estimate an encoding model for each specific subject and each part of the brain. (In fMRI the brain is divided into voxels, or volumetric units. A separate encoding model is constructed for each voxel.) Next, the encoding model is used to develop an appropriate decoding algorithm. For example, the encoding model can be inverted directly via Bayes theorem (see below). Finally, the decoding algorithm is tested on separate data set from the same observer that was not used to fit the model.

The simplest encoding model that is useful for decoding natural images is a Gabor wavelet basis model. This model describes how each voxel is tuned for space, orientation, and spatial frequency. Last year we showed that this encoding model can be used to identify, from brain activity alone, which specific image was seen by an observer, even if the image was selected at random from a database consisting of thousands of such images [7]. In fact, this approach extracted far more information from functional MRI measurements than was generally believed possible.

More recently we developed a more general spatio-temporal Gabor wavelet encoding model, and we have used this model to decode continuous time-varying natural movies from brain activity measurements. In this case we used fMRI to measure brain activity of human observers while they watched continuous, time-varying natural movies. When these models are used to perform movie identification (on a separate set of movies that were not used in fitting), we can identify which specific 20-second movie was seen by an observer with almost perfect accuracy. Furthermore, we can identify one-second movie clips to within one second of their position in the original movie. These results demonstrate that appropriate voxel-based encoding models can recover relatively fine spatio-temporal information about continuous visual experiences from brain activity measurements.

In our most recent work along these lines, we have developed a new Bayesian decoding model that can actually reconstruct natural images seen by an observer from measured brain activity. The decoder combines three elements: a structural encoding model that characterizes signals from early visual areas; a semantic encoding model that characterizes signals from higher visual areas; and appropriate priors that incorporate statistical information about the structure and semantics of natural scenes. By combining all these elements the decoder produces reconstructions that accurately reflect the distribution, structure and semantic category of the objects contained in the original image. This Bayesian decoding approach is the first practical method for reconstructing arbitrary natural images from brain activity measurements.

3 Extensions and Applications

There are several extensions of our current decoding framework that would increase the accuracy and generality of reconstructions. Our current framework uses a Gabor wavelet model to describe the structural causes of visual stimuli and a categorical model to describe their semantic causes. Alternative models that describe the higher-order statistical relationships of the structural and semantic information in natural images would produce even better reconstructions.

Our current Bayesian reconstruction algorithm operates on static, grayscale images. However, since we have also shown that it is possible to identify movies from brain activity measurements it should be possible in principle to reconstruct movies. Reconstruction of other dimensions such as color should also be possible. To the degree to which the neural mechanisms mediating subjective perceptual states such as imagery and dreams are similar to those used for normal perception, then it should also be possible to use this Bayesian framework to reconstruct subjective perceptual processes such as visual imagery and dreaming.

At this time we use fMRI to measure hemodynamic brain activity, but our framework can also be applied brain activity measurements gathered by other means: EEG, MEG, SPECT and so on. Of course, the temporal and spatial characteristics of brain activity measurements will inevitably affect the accuracy, fidelity and resolution of the decoded signals.

A perceptual brain reading device would have a wide array of potential applications. It could be used in medicine, to decode the perceptual components of hallucinations or diagnose perceptual dysfunction due to injury and disease. It could be used as a biofeedback device during rehabilitation, in order to facilitate development of neural circuits that could bypass damaged regions of the brain. More generally it would have obvious application as one component of a brain-machine interface, for rehabilitation, vehicle control and similar BMI applications.

Acknowledgments. The functional MRI scans were conducted with the support of the staff at the Brain Imaging Center at UC Berkeley and at the Veterans Administration in Martinez, CA. This research was supported by the National Eye Institute and by UC Berkeley intramural funds.

References

1. Carlson, T.A., Schrater, P., He, S.: Patterns of activity in the categorical representations of objects. J. Cog. Neurosci. 15, 704–717 (2003)
2. Cox, D.D., Savoy, R.L.: Functional magnetic resonance imaging (fMRI) 'brain reading': detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage 19, 261–270 (2003)
3. Haynes, J.D., Rees, G.: Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat. Neurosci. 8, 686–691 (2005)
4. Haynes, J.D., Rees, G.: Decoding mental states from brain activity in humans. Nat. Neurosci. Rev. 7, 523–534 (2006)
5. Hanson, S.J., Matsuka, T., Haxby, J.V.: Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a "face" area? Neuroimage 23, 156–166 (2001)
6. Kamitani, Y., Tong, F.: Decoding the visual and subjective contents of the human brain. Nat. Neurosci. 8, 679–685 (2005)
7. Kay, K.N., Naselaris, T., Prenger, R.J., Gallant, J.L.: Identifying natural images from human brain activity. Nature 452, 352–355 (2008)
8. Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J.B., Lebihan, D., Dehaene, S.: Inverse retinotopy: inferring the visual content of images from brain activation patterns. Neuroimage 33, 1104–1116 (2006)
9. Van Essen, D.C., Gallant, J.L.: Neural mechanisms of form and motion processing in the primate visual system. Neuron 13, 1–10 (1994)
10. Wu, M.C.-K., David, S.V., Gallant, J.L.: Complete functional characterization of sensory neurons by system identification. Ann. Rev. Neurosci. 29, 477–505 (2006)

Tonic Changes in EEG Power Spectra during Simulated Driving

Ruey-Song Huang, Tzyy-Ping Jung, and Scott Makeig

Swartz Center for Computational Neuroscience, Institute for Neural Computation,
University of California San Diego, La Jolla, CA, 92093-0961 USA
{rshuang,jung,scott}@sccn.ucsd.edu

Abstract. Electroencephalographic (EEG) correlates of driving performance were studied using an event-related lane-departure paradigm. High-density EEG data were analyzed using independent component analysis (ICA) and Fourier analysis. Across subjects and sessions, when reaction time to lane-departure events increased, several clusters of independent component activities in the occipital, posterior parietal, and middle temporal cortex showed tonic power increases in the delta, theta, and alpha bands. The strongest of these tonic power increases occurred in the alpha band in the occipital and parietal regions. Other independent component clusters in the somatomotor and frontal regions showed less or no significant increase in all frequency bands as RT increased. This study demonstrates additional evidence of the close and specific links between cortical brain activities (via changes in EEG spectral power) and performance (reaction time) during sustained-attention tasks. These results may also provide insights into the development of human-computer interfaces for countermeasures for drowsy driving.

Keywords: EEG, ICA, driving, alertness, delta, theta, alpha, reaction time.

1 Introduction

Drowsiness while driving is one of the major factors leading to crashes that result in severe injuries and fatalities [1-2]. Two of the basic rules of safe highway driving are to remain in the cruising lane and to keep an appropriate distance from other vehicles. Small changes in road curvature, uneven or slippery pavement, wind changes, or poor wheel alignment could make the vehicle drift out of the cruising lane. Lapses in attention and response to such lane drifts could result in collisions with other vehicles or run-off-road crashes. As drivers become fatigued and then drowsy, they exhibit slowed reaction time (RT) to traffic events, and increased deviation of vehicle lateral position (swerving) from lane center. Development of effective countermeasures to drowsy driving could prevent large numbers of serious accidents. Electro-encephalography (EEG) is one of the most direct and effective physiological measures for assessing state of arousal. Several studies have demonstrated EEG correlates of fluctuations in performance during sustained attention tasks with characteristic time scales on the order of one second to several minutes [3-15]. These studies have suggested that

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 394–403, 2009.

low-frequency EEG power, particularly in the alpha (8–12 Hz) and theta (4–7 Hz) bands, increase during periods of poor task performance (e.g., periods of high-error rate, lengthened RT, or failures to respond to driving challenges). In most studies, EEG power spectra were estimated from single-channel recordings at a few scalp sites, not allowing localization of the cortical sources of the observed EEG changes.

Our previous studies using independent component analysis (ICA) applied to high-density EEG data have demonstrated that an independent component (IC) with equivalent dipole sources located in the bilateral occipital cortex exhibits tonic changes in power spectral baseline highly correlated with performance fluctuation during sustained attention tasks, including simulated driving [12-15]. It is not known, however, whether the power spectra of other EEG processes are also strongly modulated by task performance. This study systematically explores tonic power spectral changes in the delta (1–3 Hz), theta (4–7 Hz), alpha (8–12 Hz), and beta (13–20 Hz) bands in 13 independent component clusters of brain processes obtained across subjects and sessions during periods of increased reaction time to lane-departure events during simulated driving.

2 Materials and Methods

2.1 Experimental Paradigm and Participants

A virtual-reality scene was created to simulate cruising in the fast lane of a straight highway at night. The driving simulator was implemented in C/C++ programming languages using the Open GL libraries on a desktop computer running the Linux operating system. During hour-long continuous driving sessions, computer-simulated lane-departure (deviation onset) events occurred every few seconds, during which the car drifted towards the curb or into the opposite lane with equal probability (Fig. 1). The vehicle did not ‘crash’ if the subject failed to respond but instead hit the virtual limit of the curb (after about 3 s of drift), and continued to move along the virtual curb until the subject resumed response by holding down an arrow key (response onset), and then releasing the key (response offset) when the car returned to the center of the cruising lane.

This paradigm was designed to assess subjects’ responses to perturbing events embedded in continuous monotonous driving sessions, and to monitor continuous transitions from alertness to drowsiness [12-14]. Subjects’ driving performance was measured by their reaction time (RT), defined as the duration between deviation onset and their response onset during each lane-departure event (trial). Slowed subject reaction times generally accompanied decreases in attention and alertness (Fig. 2).

Eleven right-handed healthy subjects with normal or corrected-to-normal vision participated in one or more hour-long sessions (20 sessions in all). All subjects gave informed consent before participating in an experimental protocol approved by the UCSD Human Research Protections Program. None of the subjects reported sleep deprivation the night before the experiment. Each subject had lunch about two hours before arriving at the lab around 2:00 PM; the driving experiment itself began near 3:00 PM after EEG cap and electrode set-up. During the experiment, the subject sat on a comfortable office chair with armrests 50 cm from a 19-inch monitor sitting in an EEG booth in which the background lighting was dim (~ 2–3 lux).

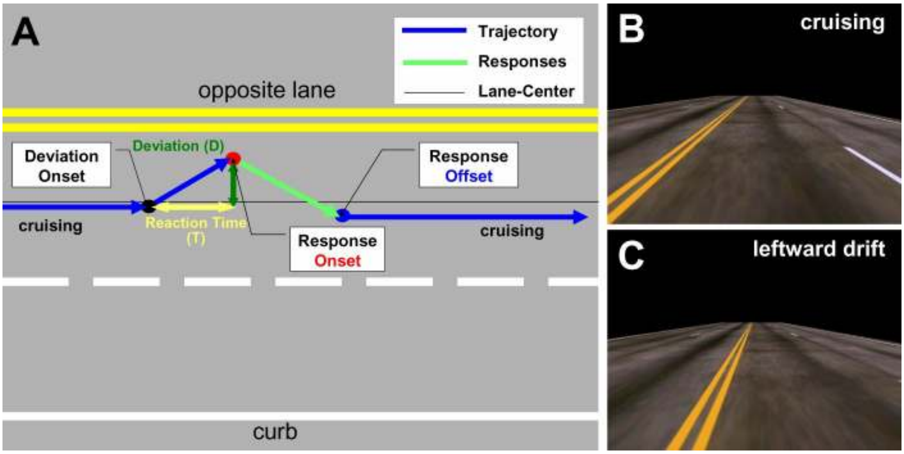


Fig. 1. Simulated driving experiment. **A.** Schematic diagram of the event-related lane departure paradigm (modified from a slide presented by R.-S.H. in [12]). **B.** A screen snapshot during cruising. **C.** A screen snapshot during a lane-departure event.

2.2 Data Acquisition and Analysis

256-channel EEG/EKG/EOG signals were recorded at 256 Hz using a BioSemi Active II acquisition system. Driving parameters (including lane positions, timing of event onsets and offsets) and subject behavioral responses were recorded at 256 Hz at the stimulus computer. A sequence of synchronized pulses was sent out from the PC parallel port to the BioSemi system for time stamping. The 3-D locations of all electrodes were digitized using a Polhemus system.

EEG data were digitally filtered using a linear FIR band pass filter (1-45 Hz) before further analysis. Continuous EEG time courses of all channels were segmented into 6-s epochs, from 1 s preceeding to 5 s following deviation onsets. Subjects typically yawned or nodded a few times during hour-long sessions. These activities caused severe artifacts across all the channels in some epochs. Channels and epochs that contained severe artifacts, including extreme values of amplitudes, large linear trends, and abnormally distributed data (high kurtosis), were rejected semi-automatically before further analysis using functions of the open source EEGLAB toolbox [16] available at http://sccn.ucsd.edu/eeglab. Channels and epochs contaminated with other sources of artifacts (blinks, eye movements, cardiac activities, and persistent head-muscle noises) were not rejected, as these artifact sources could be separated from other EEG processes using ICA described below [17-20].

The 6-s EEG epochs were concatenated into a two-dimensional matrix of size [channels, frames × epochs] after artifact rejection, and the matrix was reduced to 100 dimensions using Principle Component Analysis (PCA). Infomax ICA was applied to the dimension-reduced matrix, x, using the *binica* function with the 'extended' ICA option in EEGLAB. ICA finds an 'unmixing' matrix, W, which decomposes or linearly unmixes the matrix, x, into a sum of maximally temporally independent and spatially fixed components u, where $u = Wx$. The rows of the output data matrix, u, are time courses of activations of the independent component (IC). The ICA unmixing matrix was trained separately for each session and subject. The initial learning rate was 10^{-4}, and the training was stopped when the learning rate fell below 10^{-7}.

To test cross-subject consistency of brain processes of interest, we grouped independent components obtained from multiple sessions and subjects semi-automatically into 13 IC clusters (Figs. 3 and 4) based on their scalp maps, dipole source locations, and mean power spectral baselines [15, 21-23]. The dipole sources locations were estimated according to the digitized 3-D electrode locations and ICA weight matrix for each session using the DIPFIT2 function in EEGLAB.

For each independent component, a logarithmic power spectral baseline was computed from a 1-s window before deviation onset in each 6-s epoch extracted from its activation time course using Fast Fourier Transform (FFT). For each component cluster, the logarithmic power spectral baselines of epochs from all subjects and sessions were grouped and sorted in ascending order by trial reaction time, resulting in a matrix of size [frequency bins × epochs]. The mean logarithmic power spectra of the first 10% of epochs below 3-s RT (periods of optimal performance) were subtracted from the matrix of RT-sorted power spectral baselines at each frequency bin. The normalized matrix was further subjected to moving average across RT-sorted epochs (trials) at each frequency bin using the same window size of the first 10% epochs below 3-s RT, with a step of 10 epochs. Mean tonic power changes in delta (1–3 Hz), theta (4–7 Hz), alpha (8–12 Hz), and beta (13–20 Hz) bands were obtained from the normalized and moving-averaged power spectral matrix (Figs. 3 and 4). A two-tailed t-test was used to assess if the mean power in each moving window was statistically different from that of the first 10% epochs below 3-s RT in each frequency band using a threshold of $p<0.001$ corrected with a Bonferroni multiple comparison test.

3 Results

3.1 Driving Performance as Measured by Reaction Time

Across all sessions, subjects exhibited several fluctuations in their reaction times to lane-departure events. Some subjects became drowsy and hit the curb or drove into the opposite lane several times during hour-long sessions. Fig. 2A shows the fluctuation of subject reaction times in a representative session, where 666 lane-departure events (trials) were recorded. Fig. 2B shows the same trials sorted by reaction time (RT), which exhibit a 'bilinear' pattern (the majority of trial reaction times were short and increased exponentially in a small percentage of trials). The bilinear pattern was consistently observed across subjects and sessions, and on average 5% of trial reaction times were higher than 3 s (the approximate time lapsed before the vehicle hit the curb if the subject made no response).

3.2 Tonic Changes in Power Spectra in Relation to Reaction Time

As reaction time increased, baseline power spectra of several independent component clusters in the occipital, parietal, and temporal regions showed significant increases

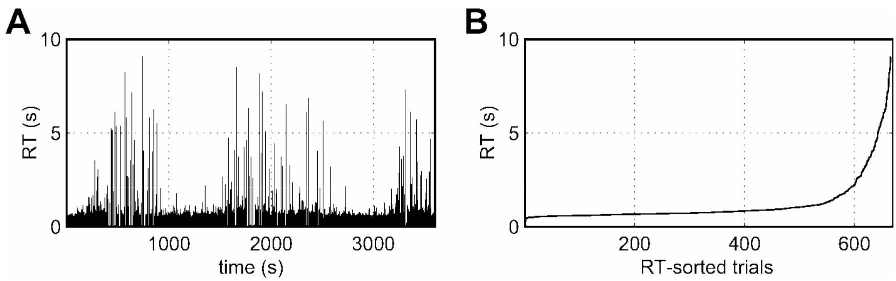


Fig. 2. Fluctuation of subject reaction times (RT) in a representative 1-hour session. **A.** Trial RTs in chronological order during the session. **B.** Trial RTs sorted in ascending order.

relative to the mean power spectra of the first 10% epochs (trials) below 3-s RT (Fig. 3). Other independent component clusters showed less significant or no tonic changes in their power spectra (Fig. 4). The observed tonic power changes in each frequency band are discussed below.

Delta Band Power (1–3 Hz). Mean baseline power in the delta band showed insignificant variations in short-RT (< 0.8 s) epochs across all IC clusters. Mean delta band power started to increase as RT increased above ~0.8 s; these tonic changes were only significant in the bilateral occipital cluster (Fig. 2A) during protracted RTs (> 2 s).

Theta Band Power (4–7 Hz). Mean baseline power in the theta band showed similar changes as in the delta band power across all IC clusters. Mean theta band power remained unchanged or even decreased slightly as RT increased moderately (< 0.8 s), and started to increase at RTs above ~0.8 s. The tonic changes were only significant during protracted RTs (near or above 3 s) in the bilateral occipital, medial posterior parietal, middle temporal, and somatomotor clusters (Figs. 3 and 4).

Alpha Band Power (8–12 Hz). Mean baseline power in the alpha band increased monotonically as RT increased in the bilateral occipital, medial posterior occipital, medial posterior parietal, and middle temporal clusters (Fig. 3). The tonic changes were significant at RTs above ~0.8 s in the occipital and parietal clusters, and were stronger than the increases in the other frequency bands. In other IC clusters, as RT increased mean alpha band power showed both insignificant increases and decreases (Fig. 4).

Beta Band Power (13–20 Hz). As RT increased, mean baseline power in the beta band showed moderate increases in the bilateral occipital, medial posterior occipital, medial posterior parietal, and middle temporal clusters (Fig. 3). The tonic changes were significant at RTs above ~0.8-s in the bilateral occipital (Fig. 3A) and medial posterior parietal (Fig. 3D) clusters. As RT increased, mean beta band power showed insignificant changes in other IC clusters (Fig. 4).

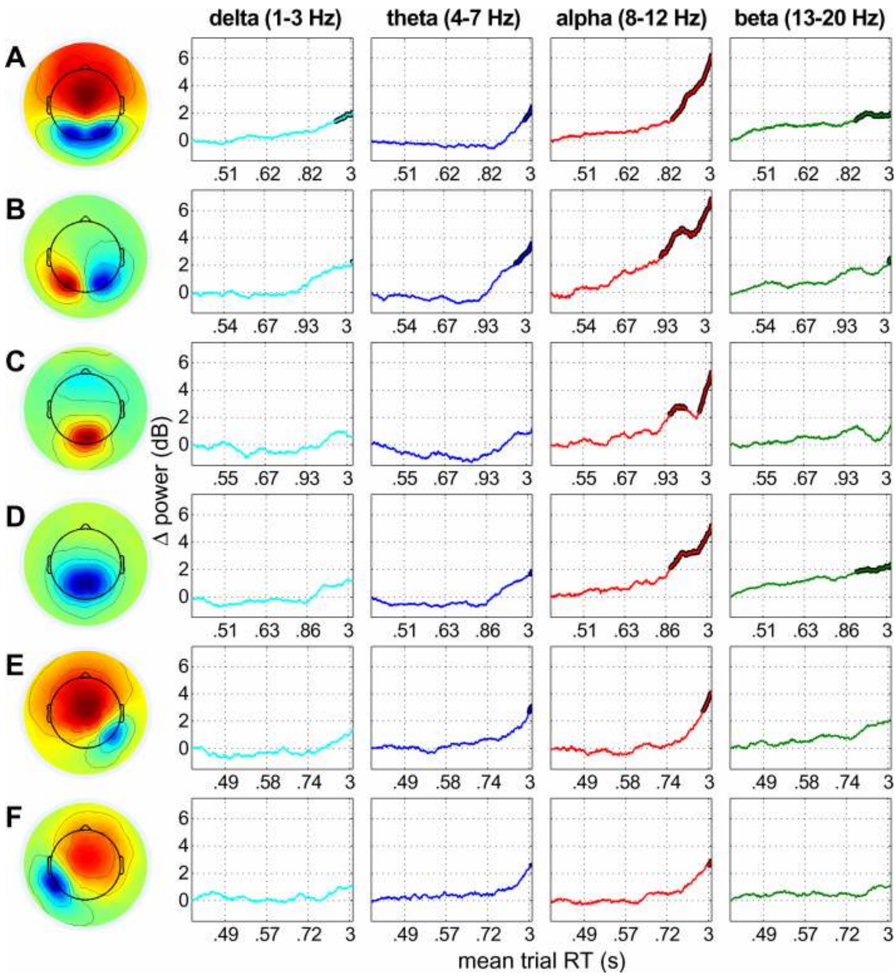


Fig. 3. Mean tonic power changes at four frequency bands for six independent component clusters. **A.** Bilaterally symmetric occipital cluster. **B.** Tangential occipital cluster. **C.** Medial posterior occipital cluster. **D.** Medial posterior parietal cluster. **E.** Right middle temporal cluster. **F.** Left middle temporal cluster. The scales of the vertical axes are the same in all subplots. The horizontal axis ticks include slow/drowsy (3-s) and 1st-3rd RT quartiles. Differences across clusters reflect differences in RT distribution in the 8–20 sessions of contributing components to each cluster. Color segments enclosed in thick black traces indicate significant ($p< 0.001$; corrected) tonic changes from the mean logarithmic power in the fastest 10% of epochs with RTs below 3-s. Note the significant changes in alpha band power in the occipital and parietal clusters.

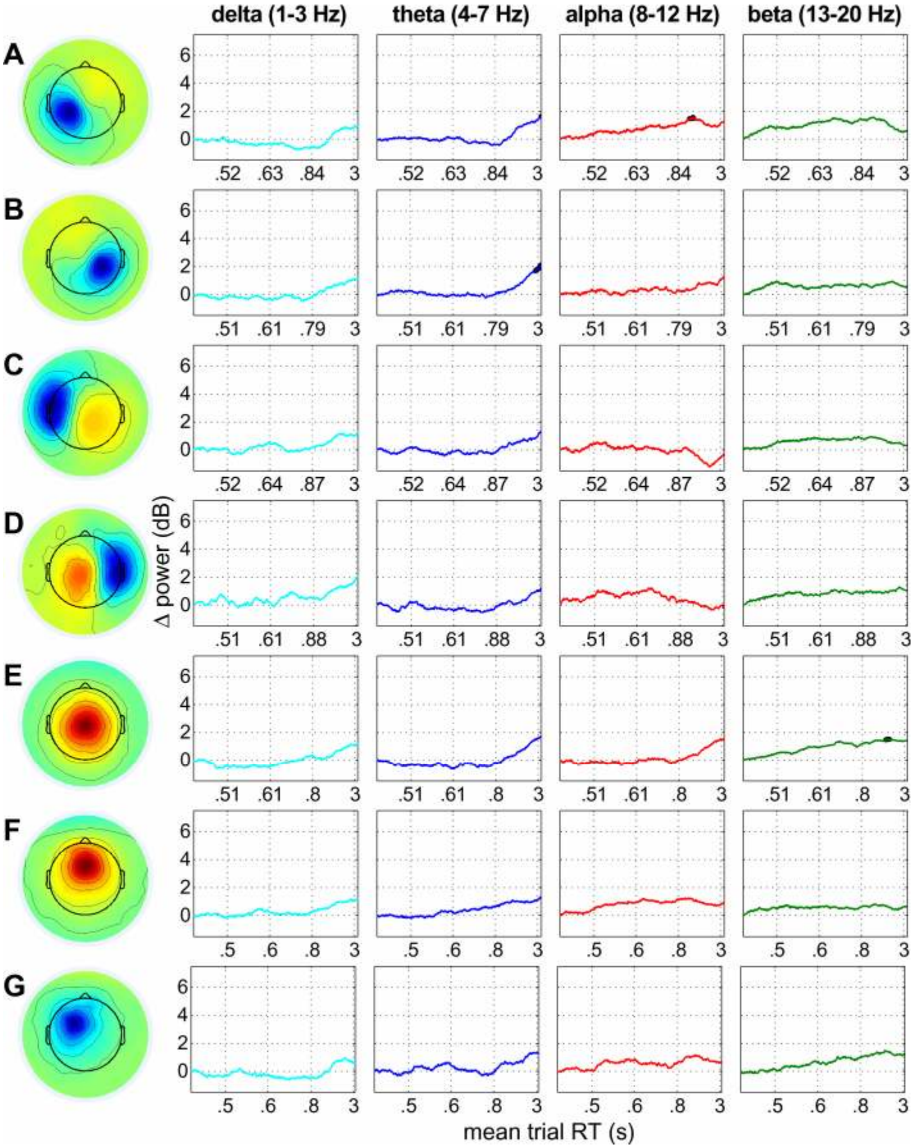


Fig. 4. Mean power in four frequency bands for seven independent component clusters showing no or weaker tonic power changes as RT increased. **A.** Left somatomotor cluster. **B.** Right somatomotor cluster. **C.** Left secondary somatosensory cluster. **D.** Right secondary somatosensory cluster. **E.** Central medial cluster. **F.** Frontal medial cluster. **G.** Left lateral frontal cluster. Other details as in Fig. 3.

4 Discussion

In this study, we applied independent component analysis to dissociate multiple brain processes whose power spectra were modulated or unaffected by performance fluctuations during simulated driving. As subject reaction time to lane-departure events increased, several clusters of independent component activities in the occipital, posterior parietal, and middle temporal cortex showed tonic power increases in the delta, theta, and alpha bands across subjects and sessions. The power spectra of other independent component clusters in the somatomotor and frontal regions were less affected or were not affected by changes in reaction time. These results provide a more comprehensive insight into brain processes involved in sustained-attention tasks.

The event-related lane departure paradigm [12-14] used in this study may provide objective and quantitative measures of both instantaneous driving performance over shorter time spans (e.g., < 10 s) and measures of average performance over longer periods (e.g., on the order of a minute). This paradigm has been replicated in other simulated driving experiments performed on a motion platform [24, 25]. Tonic power spectral changes in those experiments were similar to the results reported here; details will be reported elsewhere.

The strong tonic increases in alpha-band power in the occipital and parietal regions likely index the gradual withdrawal of visuospatial attention as drowsiness increases [26]. These increases may also be used to predict reaction time in our task and very likely during simulated driving and real-life driving. Results of this study may also help guide development of EEG-based drowsiness detection and feedback systems, and may provide useful information for evaluating systems that directly detect and apply countermeasures to drowsy driving performance, such as lane departure warning systems (LDWS) and lane keeping assistance systems (LKAS) [27-30].

Acknowledgements

This research was supported by gifts from The Swartz Foundation (Old Field, NY). This work was also sponsored in part by the U.S. Defense Advanced Research Projects Agency (DARPA) under Grant NBCH1060010 and U.S. Army Research Laboratory under Grant W911NF-08-2-0003. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the DARPA, Army, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation heron. The authors would like to thank Julie Onton, Jennifer S. Kim, and Marisa Evans for help with experimental setup, and Arnaud Delorme for assistance with data analysis.

References

1. National survey of distracted and drowsy driving attitudes and behavior, vol. 1, DOT HS 809 566 (2002), `http://www.nhtsa.dot.gov/people/injury/drowsy_driving1/survey-distractive03/index.htm`

2. Drowsy driving and automobile crashes, NCSDR/NHTSA Expert Panel on Driver Fatigue and Sleepiness,
`http://www.nhtsa.gov/people/injury/drowsy_driving1/Drowsy.html`
3. Makeig, S., Inlow, M.: Lapses in alertness: coherence of fluctuations in performance and the EEG spectrum. Electroencephalogr. Clin. Neurophysiol. 86, 23–35 (1993)
4. Makeig, S., Jung, T.P.: Changes in alertness are a principal component of variance in the EEG spectrum. NeuroReport 7, 213–216 (1995)
5. Makeig, S., Jung, T.P.: Tonic, phasic and transient EEG correlates of auditory awareness in drowsiness. Cogn. Brain Res. 4, 15–25 (1996)
6. Jung, T.P., Makeig, S., Stensmo, M., Sejnowski, T.J.: Estimating alertness from the EEG power spectrum. IEEE Trans. Biomed. Eng. 44, 60–69 (1997)
7. Makeig, S., Jung, T.P., Sejnowski, T.J.: Awareness during drowsiness: dynamics and electrophysiological correlates. Canadian J. Exp. Psy. 54, 266–273 (2000)
8. Schier, M.A.: Changes in EEG alpha power during simulated driving: a demonstration. Int. J. Psychophysiol. 37, 155–162 (2000)
9. Huang, R.S., Tsai, L.L., Kuo, C.J.: Selection of valid and reliable EEG features for predicting auditory and visual alertness levels. Proc. Natl. Sci. Counc. Repub. China B Life Sci. 25, 17–25 (2001)
10. Lal, S.K., Craig, A.: Driver fatigue: electroencephalography and psychological assessment. Psychophysiol. 39, 313–321 (2002)
11. Lal, S.K., Craig, A.: Reproducibility of the spectral components of the electroencephalogram during driver fatigue. Int. J. Psychophysiol. 55, 137–143 (2005)
12. Huang, R.S., Jung, T.P., Duann, J.R., Makeig, S., Sereno, M.I.: Imaging brain dynamics during continuous driving using independent component analysis. In: 35th Annual Meeting of the Society for Neuroscience, Washington D.C. (2005)
13. Huang, R.S., Jung, T.P., Makeig, S.: Multi-scale EEG brain dynamics during sustained attention tasks. In: Proc. IEEE ICASSP 2007, vol. IV, pp. 1173–1176 (2007)
14. Huang, R.S., Jung, T.P., Makeig, S.: Event-related brain dynamics in continuous sustained-attention tasks. In: Schmorrow, D.D., Reeves, L.M. (eds.) HCII 2007 and FAC 2007. LNCS (LNAI), vol. 4565, pp. 65–74. Springer, Heidelberg (2007)
15. Huang, R.S., Jung, T.P., Delorme, A., Makeig, S.: Tonic and phasic electroencephalographic dynamics during continuous compensatory tracking. NeuroImage 39, 1896–1909 (2008)
16. Delorme, A., Makeig, S.: EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Meth. 134, 9–21 (2004)
17. Bell, A.J., Sejnowski, T.J.: An information-maximization approach to blind separation and blind deconvolution. Neural Comput. 7, 1129–1159 (1995)
18. Lee, T.W., Girolami, M., Sejnowski, T.J.: Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. Neural Comput. 11, 417–441 (1999)
19. Jung, T.P., Humphries, C., Lee, T.W., McKeown, M.J., Iragui, V., Makeig, S., Sejnowski, T.J.: Removing electroencephalographic artifacts by blind source separation. Psychophysiol. 37, 163–178 (2000)
20. Jung, T.P., Makeig, S., McKeown, M.J., Bell, A.J., Lee, T.W., Sejnowski, T.J.: Imaging brain dynamics using independent component analysis. Proc. IEEE 89, 1107–1122 (2001)
21. Makeig, S., Westerfield, M., Jung, T.P., Enghoff, S., Townsend, J., Courchesne, E., Sejnowski, T.J.: Dynamic brain sources of visual evoked responses. Science 295, 690–694 (2002)

22. Makeig, S., Delorme, A., Westerfield, M., Jung, T.P., Townsend, J., Courchesne, E., Sejnowski, T.J.: Electroencephalographic brain dynamics following manually responded visual targets. PLoS Biol. 2, 747–762 (2004)
23. Onton, J., Westerfield, M., Townsend, J., Makeig, S.: Imaging human EEG dynamics using independent component analysis. Neurosci. Biobehav. Rev. 30, 808–822 (2006)
24. Lin, C.T., Wu, R.C., Jung, T.P., Liang, S.F., Huang, T.Y.: Estimating alertness level based on EEG spectrum analysis. EURASIP J. Appl. Signal Process. 19, 3165–3174 (2005)
25. Lin, C.T., Wu, R.C., Liang, S.F., Huang, T.Y., Chao, W.H., Chen, Y.J., Jung, T.P.: EEG-based drowsiness estimation for safety driving using independent component analysis. IEEE Trans. Circuit Sys. 52, 2726–2738 (2005)
26. Worden, M.S., Foxe, J.J., Wang, N., Simpson, G.V.: Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. J. Neurosci. 20, RC63 (2000)
27. Rimini-Doering, M., Altmueller, T., Ladstaetter, U., Rossmeier, M.: Effects of lane departure warning on drowsy drivers' performance and state in a simulator. In: Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, pp. 88–95 (2005)
28. Kozak, K., Pohl, J., Birk, W., Greenberg, J., Artz, B., Bloomer, M., Cathey, L., Curry, R.: Evaluation of lane departure warnings for drowsy drivers. In: Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, pp. 2400–2404 (2006)
29. Navarro, J., Mars, F., Hoc, J.M.: Lateral control assistance for car drivers: a comparison of motor priming and warning systems. Hum. Factors 49, 950–960 (2007)
30. Kullack, A., Ehrenpfordt, I., Lemmer, K., Eggert, F., Reflekt, A.S.: Lane departure prevention system based on behavioural control. IET Intell. Transp. Syst. 2, 285–293 (2008)

P300 Based Single Trial Independent Component Analysis on EEG Signal

Kun Li[1], Ravi Sankar[1], Yael Arbel[2], and Emanuel Donchin[2]

[1] Department of Electrical Engineering
[2] Psychology
University of South Florida, Tampa, FL, USA
li@mail.usf.edu

Abstract. A Brain Computer Interface (BCI) is a device that allows the user to communicate with the world without utilizing voluntary muscle activity (i.e., using only the electrical activity of the brain). It makes use of the well-studied observation that the brain reacts differently to different stimuli, as a function of the level of attention allotted to the stimulus stream and the specific processing triggered by the stimulus. In this article we present a single trial independent component analysis (ICA) method that is working with a BCI system proposed by Farwell and Donchin. It can dramatically reduce the signal processing time and improve the data communicating rate. This ICA method achieved 76.67% accuracy on single trial P300 response identification.

1 Introduction

As human beings, we possess a wonderful ability of communicating with other people in the world. A healthy person can express his or her ideas, feelings and desires by speech, gesturing or writing. This communicating ability makes our daily life easy and enjoyable. However, there are some people being locked in their body for different reasons. They are fully conscious and aware of what is happening in their environment but totally lose their control over any voluntary muscles. Locked-in people are not able to communicate with other persons via traditional communication method. Fortunately, with the development of neuroscience and computer science, researchers have designed a lot of different Brain Computer Interfaces to help locked-in people get their basic communicating ability back [1].

BCI is a channel established between the human brain and computer or other electronic equipments for communication and control purpose. To implement a reasonable and practicable brain computer interface there are two major prerequisites have to be fulfilled: 1. Signals that reliably describe several distinctive brain states have to be available, 2. These signals must be easily extracted and classified on-line [2]. Electroencephalography (EEG) signals meet these two prerequisites and they can be easily, noninvasively recorded, making EEG currently the best candidate for BCI system construction. There are two general types of BCI systems that have been developed by researchers using EEG as the information carriers and can be described as: Type 1: initiative BCI system [3][4][5], and Type 2: passive BCI system [6][7]. Initiative BCI system requires the subjects to learn to produce self-regulated, stable EEG signal,

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 404–410, 2009.

such as alpha or mu rhythm. This learning process will take several weeks and since there are only two states, on and off, available, it is not effective when performing multiple choices tasks. For the passive BCI system, the subjects will be given auditory or visual stimuli and generate response (Event Related Potential) to those stimuli. The Event Related Potential (ERP) can be classified after several signal processing processes and used to identify the subjects' intent. One of the well-designed passive BCI systems, known as P300 speller, was proposed by Farwell and Donchin [6] in 1988. This BCI system utilizes the P300 component of the ERP to allow locked-in individuals to communicate without using any neuromuscular function. The P300 BCI speller presents a selection of characters arranged in a 6×6 matrix. The subject focuses attention on one of the 36 character cells of the matrix in which each row or column is being intensified in a random sequence. The row and column intensifications that intersect at the attended cell represent the target stimuli, which occur with a probability of 1/6. The rare presentation of the target stimuli in the random sequence of stimuli constitutes an Oddball Paradigm [8] and will elicit a P300 response to the target stimuli. With proper P300 feature selection and classification, the attended character of the matrix can be identified and communicated [9]. This BCI system has been widely used by researchers with different signal processing techniques including Stepwise Linear Discriminant Analysis (SWLDA) [9], Support Vector Machine (SVM) [10], Matched Filter [11] and Wavelet Analysis [12]. Our research is also based on P300 speller.

Although all the techniques mentioned above demonstrated notable performance, Dean J Krusienski *et al.* [13] conclude that SWLDA is the most accurate and practical processing method on data collected using the P300 speller paradigm. However SWLDA and other techniques share the same drawback. They need to average at least several trials to remove the background noises and enhance the magnitude of P300 response before applying the P300 classifier on EEG signal. This time consuming step greatly slows down the whole signal processing and therefore makes them not suitable for the online P300 classification with single trial. We need a fast and reliable processing technique that can perform the online P300 analysis accurately for effective communication. It becomes our motivation of designing algorithms of P300 analysis based on Independent Component Analysis (ICA).

ICA is a type of blind source separation method which can break a mixed signal down to statistically independent components by maximizing their non-Gaussianity. The components are related to different features of the signal. We can map them and determine which ones are connected with P300. In other words, ICA has the ability to reveal the hidden features even if they are buried in the background noise. This ability makes it possible to detect P300 via a single trial. In this article, we discuss an ICA based single trial P300 classification algorithm that has shown 76.67% accuracy in our study.

2 Method

2.1 Data Acquisition

The subject sat upright in front of a P300 speller, focused attention on a specified letter of the matrix on the display and silently counted the number of times the target character intensified, until a new character was specified for selection. The EEG was recorded using a cap (Electro-Cap) embedded with 16 electrode locations distributed over the entire scalp. The EEG was band pass filtered 0.1–60 Hz and amplified with an amplifier (20,000×), digitized at a rate of 160 Hz.

2.2 Data Structure

The rows and columns were intensified for 75 ms with 100 ms between intensifications. Because of the delay of P300 occurrence, the EEG signal segments from 175 ms to 350 ms following each intensification are used as our experiment segments. 480 segments from each channel including 80 from target flash (the intensification of row or column that contains the desired character) and 400 from non-target flash (the intensification of row or column that does not contain the desired character) were extracted for the offline analysis.

Table 1. The EEG data structure in our experiment

Total number of EEG segments	480×16	Segment length	175 ms
Sampling Frequency	160 Hz	Number of samples in each segment	28
Intensification Duration	75 ms	Interval Time	100 ms

2.3 Preprocessing

a. All the extracted EEG signals from the 16 channels (electrodes) are low pass filtered to remove the background noise with cut-off frequency setting as 10Hz.
b. Before the independent components (ICs) of the EEG signals being computed, the observed vector x of EEG signals need to be centered and whitened to make its components uncorrelated and their variances equal unity. The whitening transformation is done by using the eigenvalue decomposition (EVD) of the covariance matrix $\varepsilon\{xx^T\}=EDE^T$, where E is the orthogonal matrix of eigenvectors of $\varepsilon\{xx^T\}$ and D is the diagonal matrix of its eigenvalues. The whitening can now be expressed as:

$$\tilde{x} = ED^{-1/2}E^T x \tag{1}$$

If we express x as:

$$x = As \tag{2}$$

where s is the independent components vector and A is the linear transformation from s to x, then we have:

$$\tilde{x} = ED^{-1/2}E^T As = \tilde{A}s \tag{3}$$

It can be easily verified that the new transformation matrix $\tilde{A}$ is orthogonal. Hence the number of parameters needs to be estimated reduced from n^2 (the number of elements in the original matrix A) to about n(n-1)/2 ($\tilde{A}$ contains only n(n-1)/2 degree of freedom) [14].

2.4 Independent Component Analysis

Independent component analysis (ICA) is a statistical and computational technique for revealing hidden factors that underlie sets of random variables, measurements, or signals. It is a good solution for the Blind Source Separation (BSS) problem. For example, two speakers (S_1 and S_2) speak simultaneously in a room with two recorders (R_1 and R_2) recording their speech at different location in the room. The recorded signals, $R_1(t)$ and $R_2(t)$, can be expressed like this:

$$R_1(t) = a_{11}S_1(t) + a_{12}S_2(t)$$
$$R_2(t) = a_{21}S_1(t) + a_{22}S_2(t) \tag{4}$$

If we know the values of a_{11}, a_{12}, a_{21} and a_{22}, we can solve these equations for S_1 and S_2. Unfortunately these weights (a's) are unknowns and these equations can only be solved under the assumption that S_1 and S_2 are independent non-Gaussian signals by Independent Component Analysis. This is a famous example of "cocktail party" problem. Obviously, EEG signal analysis is a type of "cocktail party" problem. The electrodes "record" the mixed EEG signal at different locations around the scalp. Therefore, it is reasonable to apply ICA on EEG signal to identify those independent sources and map them to P300.

There are a lot of ICA algorithms available, such as Infomax[15], JADE[16] and FastICA[17]. All of them can successfully compute the independent components by maximizing the non-Gaussianity or *negentropy*, which is a measurement of non-Gaussianity [18], of the ICs. In our research, we choose FastICA to perform ICA because it converges much faster than other algorithms with high reliability.

We use the average of 400 of preprocessed 175ms EEG signals from non-target flash as the "standard non-target flash" signal, denoted as x_{nt}. Similarly, the average of 80 preprocessed EEG signals from target flash is set as the "standard target flash" signal, denoted as x_t. By applying FastICA, the independents components vector s and the mixing matrix A of x_t can be computed and expressed as:

$$x_t = A_t s_t \tag{5}$$

The vectors in s_t are used as the "standard independent components set" of the EEG signal and A_t is used as the "standard coefficients matrix" showing the activation status of the ICs underlying in x_t. Here we made an assumption that the EEG signal from target flash contains more components than those from non-target flash. This is reasonable since the EEG signal of target flash is constituted of "background noise" and P300 response while the EEG signal of non-target flash is constituted of "background noise" only. By substituting s_t and x_{nt} in equation (2), we can solve for A_{nt} that shows the activation status of the ICs underlying in x_{nt}. We inspected A_{nt} and A_t and noticed the significant differences between them. The coefficients of some ICs are positive or negative in A_t while they have opposite sign in A_{nt}. It implies that some of the ICs in the standard ICs set are strongly related to P300 response. In order to confirm this conclusion, we computed the activation matrix A of the 400/80 non-target/target flashes individually. By inspection, we find out that most of them somehow follow the “sign rule” mentioned before. This fact inspires us to use the sign of the coefficients of the ICs as the feature for P300 identification.

2.5 Majority Vote Scenario

There are 5 ICs that are considered to have strong relationship to P300 component can be used to determine whether an incoming EEG signal is from a target flash. However each of these ICs comes from 16 different channels in which their coefficient sign may vary. Therefore we first set up the majority voting rule to determine the vote that an IC will give. The 16 coefficients of an IC in different channels form a column of the activation matrix A. To make sure there is a “majority” voting, we drop the most unstable channel and set the rule as:

$$\text{If } \frac{a_{ij}}{a^*_{ij}} > 0 \quad v_i = 1\text{; otherwise } v_i = -1 \tag{6}$$

Where a_{ij} is an element of the activation matrix A at the *i-th* row and the *j-th* column; a^*_{ij} is an element in activation matrix A_t at the *i-th* row and the *j-th* column; j is the index of specific voting IC. Apparently, if a_{ij} and a^*_{ij} have the same sign, $v_i = 1$; if they have opposite sign, $v_i = -1$. Therefore, the chosen ICs make their votes according to the following rule:

$$\text{if } \sum_{i=1}^{n} v_i > c \quad \text{(n=15), vote for “Target”, otherwise vote for “Non-target”} \tag{7}$$

where c is an integer that can be chosen from 0 to 15 to control the type I (target identified as non-target) and type II (non-target identified as target) error. In our research, we set $c = 2$. The majority votes of the chosen ICs determine the label of an incoming EEG signal.

2.6 Processing Flow

The processing flow used in this work is given in Fig 1.

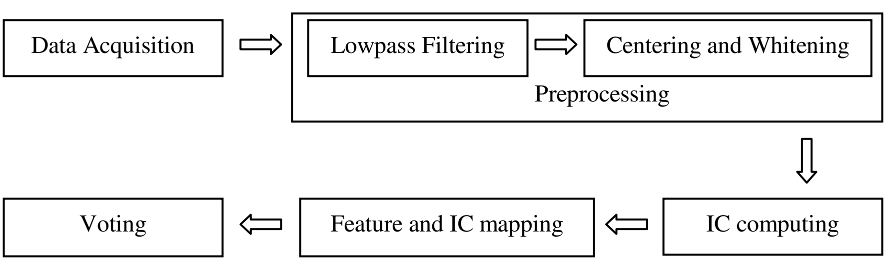


Fig. 1. The processing flow

3 Results and Discussion

In this work, we choose 3 or 5 P300 related ICs as the voters and classify the incoming EEG signals according to their majority votes. This scenario is tested by 180 EEG signals including 150 from non-target flash and 30 from target flash. When 3 ICs are employed, 60% of the incoming signals are correctly identified. When 5 ICs are employed, 76.67% of them are correctly identified. In our research we prefer to reduce the type II error because if we fail to identify a target flash, the identification process can be repeated till the target successfully identified. But if a signal is falsely identified as "Target", this error will not be realized until the final character selection. Considering this, we may reduce the type II error by increasing the c value. However, the tradeoff is that the processing time will increase due to repetition. The P300 based single trial ICA algorithm significantly reduces the processing time by removing the time consuming step due to "averaging" used in other algorithms. Furthermore, our algorithm will stop and start the next "Target searching" whenever it hits a "Target". Thus the expecting target identifying time is given by $\varepsilon(t) = 3.5 flashes = 175 \times 3.5 = 612.5\ ms,$ which is approximately 1/10 of the best processing time achieved by SWLDA [19].

There is still room for improving the processing speed and accuracy by optimizing the algorithm. For example, we can weigh the voters or modify the voting rule to improve the performance of voting. In our experiment, we made an assumption that the P300 response occurs between 175 ms and 350 ms following a target flash, which is not true for some subjects because in some cases P300 shows up in the 350 ms to 500 ms range. This problem can be solved by using appropriate flashing and interval time. We are planning to optimize our algorithm by applying appropriate filter during preprocessing, solving the non-stationary problem [20] and involving statistical models in our future work. Our goal is to further improve the accuracy of the single trial P300 analysis algorithm to make it more suitable for real-world application and clinical use.

References

1. Hoffmann, U., Vesin, J.M., Ebrahimi, T.: Recent Advances in Brain-Computer interfaces. In: IEEE International Workshop on Multimedia Signal Processing (MMSP 2007) (2007)
2. Kalcher, J., Flotzinger, D., Neuper, C., Gölly, S., Pfurtscheller, G.: Graz brain-computer interface II: towards communication between humans and computers based on online classification of three different EEG patterns. Med. & Biol. Eng. & Comput. 34, 382–388 (1996)
3. Wolpaw, J.R., McFarland, D.J., Neat, G.W., Forneris, C.A.: An EEG-based brain-computer interface for cursor control. Electroencephalography and Clinical Neurophysiology 78, 252–259 (1991)
4. McFarland, D.J., Neat, G.W., Read, R.F., Wolpaw, J.R.: An EEG-based method for graded cursor control. Psychobiology 21(1), 77–81 (1993)
5. Pfurtscheller, G., Flotzinger, D., Kalcher, J.: Brain-Computer Interface – a new communication device for handicapped persons. Journal of Microcomputer Applications 16, 293–299 (1993)

6. Farwell, L.A., Donchin, E.: Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalography and Clinical Neurophysiology 70, 510–523 (1988)
7. Rebsamen, B., Burdet, E., Guan, C.T., Zhang, H.H., Teo, C.L., Zeng, Q., Ang, M., Laugier, C.: A Brain-Controlled Wheelchair Based on P300 and Path Guidance. In: Proceedings of IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics (2006)
8. Fabiani, M., Gratton, G., Karis, D., Donchin, E.: Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. Adv. Psychophysiol. 2, 1–78 (1987)
9. Krusienski, D.J., Sellers, E.W., McFarland, D.J., Vaughan, T.M., Wolpaw, J.R.: Toward enhanced P300 speller performance. J. Neurosci. Methods 167(1), 15–21 (2008)
10. Kaper, M., Meinicke, P., Grossekathoefer, U., Lingner, T., Ritter, H.: BCI Competition 2003—Data Set IIb: Support Vector Machines for the P300 Speller Paradigm. IEEE Transactions on Biomedical Engineering 51(6) (June 2004)
11. Serby, H., Yom-Tov, E., Inbar, G.F.: An Improved P300-Based Brain-Computer Interface. IEEE Transactions on Neural Systems and Rehabilitation Engineering 13(1) (March 2005)
12. Bostanov, V.: BCI Competition 2003–Data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram. IEEE Trans Biomed. Eng. 51, 1057–1061 (2004)
13. Krusienski, D.J., Sellers, E.W., Cabestaing, F., Bayoudh, S., McFarland, D.J., Vaughan, T.M., Wolpaw, J.R.: A comparison of classification techniques for the P300 speller. J. Neural Eng. 3, 299–305 (2006)
14. Hyvärinen, A.: Fast ICA by a fixed-point algorithm that maximizes non-Gaussianity. Independent Component Analysis Principles and Practice. Cambridge University Press, Cambridge (2001)
15. Bell, A.J., Sejnowski, T.J.: An information-maximisation approach to blind separation and blind deconvolution. Neural Computation 7, 1129–1159 (1995)
16. Cardoso, J.F., Comon, P.: Equivariant adaptive source separation. IEEE Transactions on Signal Processing 45(2), 434–444 (1996)
17. Hyvärinen, A.: Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. IEEE Transactions on Neural Networks 10(3), 626–634 (1999)
18. Hyvärinen, A., Oja, E.: Independent Component Analysis: Algorithms and Applications. Neural Networks 13(4-5), 411–430 (2000)
19. Donchin, E., Spencer, K.M., Wijesinghe, R.: The Mental Prosthesis: Assessing the Speed of a P300-Based Brain–Computer Interface. IEEE Transactions on Rehabilitation Engineering 8(2) (June 2000)
20. Penny, W.D., Roberts, S.J., Curran, E.A., Stokes, M.J.: EEG-based communication: A pattern recognition approach. IEEE Trans. Rehab. Eng. 8, 214–215 (2000)

Directed Components Analysis: An Analytic Method for the Removal of Biophysical Artifacts from EEG Data

Phan Luu[1], Robert Frank[2], Scott Kerick[3], and Don M. Tucker[1]

[1] Electrical Geodesics, Inc., Eugene, Oregon
[2] NeuroInformatics Center, University of Oregon, Eugene, OR
[3] US Army Research Laboratory, Aberdeen Proving Ground, MD 21005
pluu@egi.com, rmfrank@mac.com, skerick@arl.army.mil,
dtucker@egi.com

Abstract. Artifacts generated by biophysical sources (such as muscles, eyes, and heart) often hamper the use of EEG for the study of brain functions in basic research and applied settings. These artifacts share frequency overlap with the EEG, making frequency filtering inappropriate for their removal. Spatial decomposition methods, such as principal and independent components analysis, have been employed for the removal of the artifacts from the EEG. However, these methods have limitations that prevent their use in operational environments that require real-time analysis. We have introduced a directed components analysis (DCA) that employs a spatial template to direct the selection of target artifacts. This method is computationally efficient, allowing it to be employed in real-world applications. In this paper, we evaluate the effect of spatial undersampling of the scalp potential field on the ability of DCA to remove blink artifacts.

Keywords: EEG, artifact, brain activity, neuroergonomic.

1 Introduction

The electroencephalogram (EEG) is a scalp measure of electrical currents produced by cortical neurons. The EEG is the oldest neuroimaging technique in use today for both basic science research and clinical diagnosis. With the development of dense-sensor arrays (greater than 128-channels) and accurate head models that describe the propagation of current from the cortex to the scalp, it is now possible to study the brain with excellent spatial (on the order of cm) as well as temporal resolution. This is particularly important as researchers and clinicians are interested in both kinds of information. These same strengths, as well as its low cost and relative portability, make EEG technology ideally suited for neuroergonomic applications.

However, because the EEG is a measure of electrical potentials, it can be easily contaminated with artifacts from sources that also generate current. These artifacts may interfere with visual interpretations of the EEG as well as analysis. It is now well known that noise (i.e., artifacts) contributes significantly to source estimate errors [1]. Particularly serious are artifacts of biological origin, such as from the heart or eyes,

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 411–416, 2009.

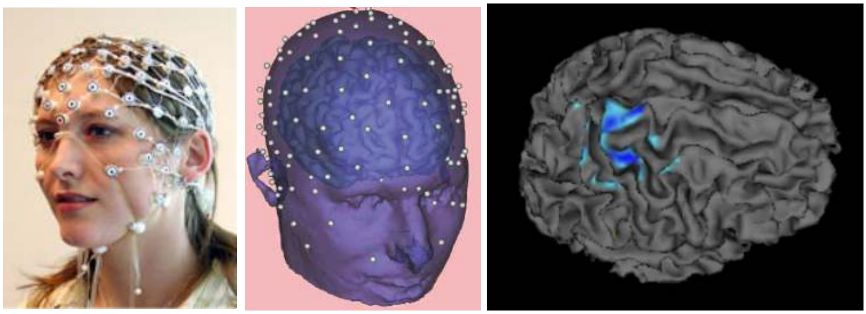


Fig. 1. Left: Subject with 128-channel Hydrocel Sensor Net. Middle: Head model with sensors registered to scalp surface. Right: Cortical surface with current source estimate.

because they may be differentially distributed across experimental conditions, such as less blinks during conditions of high workload. If these artifacts are not removed, they may be perfectly confounded with the experimental conditions.

Because they contaminate the energy spectrum of the EEG, these biophysical artifacts cannot be removed by simple time-series filtering techniques. Therefore, common approaches to dealing with these artifacts involve construction of spatial filters through different statistical decomposition techniques, such as independent components analysis (ICA) or principle components analysis (PCA). These approaches require the user to manually identify the component(s) that captures the blink activity and remove them during data reconstruction. There are several limitations to these approaches. First, the spatial distribution of blinks is often described by multiple components, making the process of their removal more complicated and arduous. Second, these methods tend to be computationally intensive (such as ICA with the Infomax technique) and cannot be used for real-time applications.

Directed Components Analysis (DCA) is an ongoing development of a modular, computationally efficient, EEG artifact extraction algorithm [2], based upon the work of Ille et al. [3]. DCA segments multichannel EEG data into a sequence of overlapping windows, uniquely modeling each window's artifactual and cortical activity with topographies derived from the data itself. This segmenting approach enables DCA to accommodate the temporal evolution of the artifactual and cortical activities that constitute the scalp recorded EEG, from which segment-specific artifactual and cortical models are then derived. The segment-specific spatial filters extract any and all activity that is both spatially correlated to the EEG segment's artifactual model topographies and not described by its cortical model topographies. This sharing of variance between artifactual and cortical models prevents extraction of cortical activity spatially correlated to the artifacts of interest.

Since DCA, like ICA and PCA, creates spatial filters to remove artifacts, its performance is dependent upon accurate description of the spatial topography of the brain signal as well as the artifact. Like discretization of a time series, wherein the Nyquist theorem states that the sampling (i.e., discretization) rate must be two times that of the highest frequency to avoid aliasing of energy from the highest frequency into lower frequency spectra, accurate discretization of spatial information requires adequate spatial sampling; if not adequately sampled, high spatial frequency information will be smeared and mistaken for low-spatial frequency information [4].

In this paper, we compare the effects of undersampling the scalp potential field on the ability of DCA to remove blink artifacts with minimal distortion of the underlying EEG. We start with a 128-channel EEG data set [5] and subsample the data down to 32-channels. We then applied DCA to both data sets and compared the integrity of these two data sets after artifact cleaning.

2 Method

The approach that we take to the evaluation of DCA's ability to clean data sets of differing spatial resolution involves the use of event-related potentials (ERPs) as the target signal. We evaluated whether DCA would significantly distort the ERP. An ERP is an average of single trials of EEG locked to an event of interest (such as target, feedback, or motor response onsets). An assumption of ERP averaging is that brain responses elicited by an event of interest is approximately constant from trial to trial. Through averaging, the background EEG will be attenuated (since its amplitude is random) and the ERPs will be accentuated.

The data set employed for this study is described elsewhere [5,6]. The continuous EEG record was segmented relative to the onset of the target stimulus (-200 ms before to 1000 ms after). Automated detection of blink artifacts was performed on each segment (i.e., trial or epoch) of the EEG. An EEG expert (one of the authors) reviewed the automated artifact detection results to ensure accuracy of the artifact detection. The segments were then classified as either being contaminated with blinks or being blink-free; this constitutes files of class 1. Next, the original continuous EEG file was submitted to DCA.

To extract eye blinks, DCA first constructs segment-specific spatial filters based upon the corresponding segment's eye blink (i.e., the EEG segment that contains a blink artifact) and cortical models (EEG segment that does not contain a blink artifact). DCA eye-blink spatial filters extract the intensity of any activity correlated to their segment's blink topography that is not in the span of the segment's cortical model eigenvectors. The degree of the extracted intensity is in proportion to the amount of correlation, and is computed, at each time point, as the inner product of the spatial filter and the scalp-recorded EEG. The result is a time course of eye blink intensity, which, when multiplied by its blink topography, estimates the eye blink potential on each channel at each segment time point.

DCA spatial filters distinguish between eye blinks and spatially correlated cortical activity by virtue of their construction: each segment's spatial filter is a weighted sum of the segment's left-nullspace eigenvectors. Left-nullspace eigenvectors are the covariance-matrix-derived eigenvectors that are not utilized by the segment's cortical model. As the totality of the covariance-matrix-derived eigenvectors span all of channel space, linear combinations of the left-nullspace eigenvectors model that portion of any scalp topography that cannot be otherwise expressed as a linear combination of the cortical model eigenvectors. DCA derived spatial filters therefore estimate the time course of eye blink intensity based solely upon their correlation to that portion of the scalp recorded EEG in the span of the left nullspace. The DCA identification of eye-blink contaminated intervals, and the exclusion of their topographies from the cortical model, ensures to the extent possible that the artifactual topographies of interest, eye blinks, are the primary component of the EEG in the left-nullspace's span.

DCA optimizes spatial filter performance through a technique we term left-nullspace filtering.

1. A cortical model is first composed of eigenvectors that account for ~ 95% of the blink-free EEG variance in the current windowed segment. The exact percentage is specified by the user, though 95% is typical.
2. A Moore-Penrose pseudo-inverse applied to an augmented matrix of the eye blink typography and cortical model eigenvectors generates the spatial filter as a weighted sum of left-nullspace eigenvectors.
3. Nullspace filtering deletes from the left nullspace all eigenvectors with blink topography correlation less than a specified threshold, typically 50% of the maximum value.
4. The deleted left-nullspace eigenvectors from step 3) are moved into the cortical model to augment its span.
5. The final spatial filter is regenerated as a weighted sum of the remaining left null-space eigenvectors.
6. The inner product of this final spatial filter with the EEG electrode potentials at each segment time point extracts the time course of intensity of the segment's blink topography.

After removal of blinks, the file was segmented and submitted to the automated artifact detection. This constitutes files of class 2. In the class 2 files, the ERP was derived from only those segments that were originally classified as containing blinks, from the class 1 files, but are now cleaned of the artifacts (using the same number of trials that went into the derivation of the ERP of blink-free data, see below). We then evaluated whether this ERP derived from blink-cleaned data contained any statistically significant distortions. This was done by estimating error bounds, obtained by creating a distribution of ERP of the blink-free segments from the class 1 files.

The distribution of the blink-free ERP was derived using the following procedure. First, from the total blink-free pool we took random subsets of 25% of the trials to derive each ERP (the ERP from all files contained no fewer then 25 segments). This ensured that as we obtained multiple samples of the ERPs that there would be enough variability of the ERP estimates to produce an accurate characterization of the variability of data points in this distribution. The distribution of the ERPs for each file contained 1,000 observations (i.e., averages). Second, from each distribution we identified the upper and lower 2.5 percentile.

After these steps are performed, for each EEG channel, we have an ERP derived from blink-cleaned data with error bounds to indicate whether the ERP (at each sample and each EEG channel) has been distorted by the cleaning procedure (see Figure 2). This procedure was performed on 12 data files (six 128-channel and six 32-channel files).

We then computed an error fraction to summarize the results for each data file. The error fraction was defined as the fraction of samples in the blink-cleaned data that exceeded the error bounds. An error fraction was computed for each channel and then averaged over all channels to produce one summary.

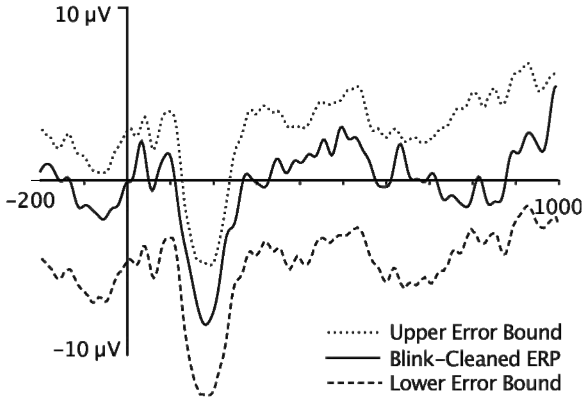


Fig. 2. Blink-cleaned ERP with error bound estimates from an EEG channel over the left occipital scalp region (averaged over 6 files from the 128-channel data set)

3 Results

The results are presented in Table 1. The average error fraction for the 128-channel data is .22 and .34 for the 32-channel data, indicating that ERPs derived from lower-channel counts are more distorted by DCA. Examination of the data showed that for files 2 and 3 (128-channel) the data tend to be distorted in those EEG channels located at the front of the head (by the eyes). When we examined the 32-channel data set, all files demonstrated significant distortion of the ERP in those EEG channels over the front of the head.

Table 1. Error fractions by channel-count

	128-Channels	*32-Channels*
	Mean (SD)	
1	.17 (.12)	.37 (.26)
2	.27 (.18)	.29 (.31)
3	.30 (.20)	.35 (.32)
4	.21 (.14)	.37 (.27)
5	.18 (.15)	.20 (.18)
6	.17 (.12)	.47 (.30)

4 Discussion

This preliminary study showed that the ability to remove blinks from the EEG, using DCA (which is a spatial filtering method) is dependent on the accurate characterization of the spatial distribution of the blinks. Across the 12 data files, the ERP data were consistently more distorted for the 32-channel compared to the 128-channel data sets. The spatial distribution of blinks is poorly captured by the 32-channel data set and correspondingly, the cleaned ERP data showed more distortion.

One question is whether higher-channel data (such as 256-channels) will show less distortion. We will examine this in future studies. Related to this question, and relevant to in-field EEG applications, is whether DCA can be made to work with sparse-array channel counts while minimizing distortion. In-field applications often employ lower channel counts (often 3 channels). These lower channel counts systems are believed to be more acceptable to the end-user because they are perceived as being less obtrusive. Data from lower channel counts, obviously, cannot be employed to estimate sources of the EEG because they do not accurately describe the scalp distribution of the EEG. With the objective of cleaning artifacts using spatial filtering techniques, it is clear that adequate spatial sampling is still required. Therefore, to minimize distortions of the EEG with lower channel count configurations, it is unavoidable that more sensors are required. However, it may be possible that, for the purpose of cleaning eye-related artifacts, additional sensors (perhaps 12 or less) placed around the eyes, forehead, and face areas may suffice because it is the artifact topography that needs to be accurately characterized. However, how this may bias creation of the filters, because the brain signals are now less specified, will need to be determined.

Finally, we are unaware of any study that has employed the approach (permutation statistics) we used in the present work for evaluation of the effectiveness of ICA or PCA for the removal of artifacts. Thus, it would be informative for future studies to compare DCA against these other methods using the same evaluative framework.

References

1. Cuffin, B.N., Schomer, D.L., Ives, J.R., Blume, H.: Experimental Tests of EEG Source Localization Accuracy in Spherical Head Models. Clin. Neurophysiol. 112, 46–51 (2001)
2. Poolman, P., Frank, R.M., Bell, R.M., Tucker, D.M., Luu, P.: Advanced Integrated Real-Time Artifact Removal Framework. In: Schmorrow, D.D., Stanney, K.M., Reeves, L.M. (eds.) Foundations of Augmented Cognition, pp. 102–110. Falcon Books, San Ramon (2006)
3. Ille, N., Berg, P., Scherg, M.: Artifact Correction of the Ongoing EEG Using Spatial Filters Based on Artifact and Brain Signal Topographies. J. Clin. Neurophysiol. 19, 113–124 (2002)
4. Srinivasan, R., Tucker, D.M., Murias, M.: Estimating the Spatial Nyquist of the Human EEG. Behav. Res. Meth. Inst., Comput. 30, 8–19 (1998)
5. Luu, P., Shane, M., Pratt, N.L., Tucker, D.M.: Corticolimbic Mechanisms in the Control of Trial and Error Learning. Brain Res. 1247, 100–113 (2009)
6. Luu, P., Tucker, D.M., Stripling, R.: Neural Mechanisms for Learning Actions in Context. Brain Res. 1179, 89–105 (2007)

Functional Near-Infrared Spectroscopy and Electroencephalography: A Multimodal Imaging Approach

Anna C. Merzagora[1], Meltem Izzetoglu[1], Robi Polikar[1,2], Valerie Weisser[3], Banu Onaral[1], and Maria T. Schultheis[1,3]

[1] School of Biomedical Engineering, Science and Health Systems, Drexel University, 3141 Chestnut Street, Philadelphia PA 19104, U.S.A.
[2] Department of Electrical and Computer Engineering, Rowan University, 201 Mullica Hill Road, Glassboro NJ 08028, U.S.A.
[3] Department of Psychology, Drexel University, 3141 Chestnut Street, Philadelphia PA 19104, U.S.A.

{a.merzagora,meltem.alkan.izzetoglu}@drexel.edu, polikar@rowan.edu, {vdw25,banu.onaral,schultheis}@drexel.edu

Abstract. Although neuroimaging has greatly expanded our knowledge about the brain-behavior relation, combining multiple neuroimaging modalities with complementing strengths can overcome some limitations encountered when using a single modality. Valuable candidates for a multimodal approach are functional near-infrared spectroscopy (fNIRS) and electroencephalography (EEG). fNIRS is an imaging technology that localizes hemodynamic changes within the cortex. However, hemodynamic activation is an intrinsically slow process. On the other hand, EEG has excellent time resolution by directly measuring the manifestation of the brain electrical activity at the scalp. Based on their complementary strengths, the integration of fNIRS and EEG may provide higher spatiotemporal resolution than either method alone. In this effort, we integrate fNIRS and EEG to evaluate the behavioral performance of six healthy adults in a working memory task. To this end, features extracted from fNIRS and EEG were used separately, as well as in combination, and their performances were compared against each other.

Keywords: multimodal neuroimaging, functional near-infrared spectroscopy, EEG, pattern classification, working memory, n-back, P300.

1 Introduction

The last 20 years have seen a rapid advance in neuroimaging technologies that are now widely used for non-invasive investigation of human brain functions. Application of these technologies to the fields of basic and clinical neuroscience has greatly expanded our knowledge about brain activity associated with perceptual, cognitive, emotional and behavioral processes, in health [1],[2] and disease [3],[4]. In particular, neuroimaging techniques have contributed to the investigation of the specialization

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 417–426, 2009.

and integration of different cerebral areas in the normal brain and to the study of brain dysfunction in varying disorders [5],[6]. Nonetheless, the current understanding of the relation between brain activity and behavior is still limited. One of the restricting factors is the inherent complexity of the system to be investigated. In fact, most task designs in neuroimaging aim at probing or manipulating one cognitive domain at a time, but human behavior results from the interaction of multiple components (e.g., attention, orienting response, planning or short-term memory). Additionally, the macroscopic brain activity is a multifaceted process and the combined use of multiple neuroimaging technologies could capture different aspects of this process.

On one hand, techniques such as electroencephalography (EEG) and magnetoencephalography (MEG) record the integrated and synchronized electromagnetic activity of populations of pyramidal neurons in the cerebral cortex [7],[8]. Both EEG and MEG have excellent temporal resolution (at the millisecond level), but they also share the weakness of a poor three-dimensional spatial localization: the activated cortical sources need to be estimated based on the distribution of the electromagnetic fields on the scalp ("inverse problem"), which is a mathematically ill-posed problem.

On the other hand, more recent neuroimaging technologies focus on hemodynamic changes, an indirect measure of brain function. These changes consist in variations in regional blood flow, in blood oxygenation or in local metabolism and are generally assumed to reflect changes in the neural activity [9]. The neuroimaging modalities in this group are functional magnetic resonance imaging (fMRI), functional near-infrared spectroscopy (fNIRS), positron emission tomography (PET) and single photon emission computed tomography (SPECT). In contrast to EEG and MEG, the hemodynamic-based technologies offer the advantage of providing information about the spatial location of the recorded activity (with a resolution down to few millimeters). However, hemodynamic changes are intrinsically slow processes, happening in the range of seconds [10], thus limiting the temporal resolution of the recordings.

Therefore, given the complexity of the investigated processes and the wide range of characteristics for the different imaging technologies, the use of multimodal approaches is gaining the interest of the scientific community [11],[12],[13],[14]. The underlying principle is that all neuroimaging techniques provide in vivo measures of brain function but each has its own set of assets and drawbacks. Hence, the combination of multiple imaging modalities with complementing strengths can partially overcome the limitations encountered by each individual modality.

Two valuable candidates for a multimodal approach are fNIRS and EEG. fNIRS is a brain imaging technology that relies on optical techniques to detect changes in the hemodynamic activity within the cortex in response to sensory, motor, or cognitive activation [15],[16]. fNIRS relies on the placement of near-infrared light sources and detectors on the scalp. Oxygenated (HbO2) and deoxygenated (HHb) hemoglobin are the dominant light absorbing elements within the brain at the near infrared wavelengths and have different absorption patterns of light. Thus, fNIRS can record changes in HbO2 and HHb concentrations, which occur during brain activation [17]. Similar to other hemodynamic-based neuroimaging modalities, fNIRS is able to provide information about the specific localization of the recorded hemodynamic activity. Compared to fMRI or PET, however, fNIRS is affordable and easily implementable in a portable system, allowing for a wider range of applications. By associating fNIRS with EEG, we can additionally take advantage of the good temporal resolution provided by the latter. After all, EEG can capture information about rapid cortico-cortical or thalamocortical oscillations that play a crucial role in the elaboration and integration of information in cognitive networks. Therefore, based on the complementary strengths offered by EEG and fNIRS, their integration has the potential to provide higher spatio-temporal resolution than either method alone.

2 Example of Multimodal Approach

2.1 Background

We provide here an example of multimodal imaging approach using fNIRS and EEG. The aim is evaluate the performance level of subjects during a task with high working memory load. This was pursued by using measures from fNIRS or EEG individually or in a combination, the results of which were then compared. The rationale for using these modalities among others is two-fold. First, there are indications that the oxygenation changes recorded by fNIRS in working memory tasks are related to the task load and at the same time are affected by the performance level of the subject [18]. Second, EEG has been extensively applied in working memory research. In particular, many studies focused on the P300 component, a peak occurring about 300 ms after a target stimulus presentation and reflecting the demand on attentional resources [19]. Based on its neuropsychological interpretation, the P300 amplitude is expected to increase with increasing task complexity [20], but studies have shown a decline when the stimulus is objectively harder to discriminate or when the subject is less confident in its discrimination [21]. Hence, the combined use of fNIRS and EEG can provide insight into the different mechanisms underlying the observed low performance on a working memory task.

2.2 Experimental Protocol

Six subjects (3 males and 3 females) were selected from a larger pool of healthy participants. All subjects were right-handed, with vision correctible to 20/20. Participants denied any history of neurological disorders, psychiatric illness, substance abuse or being on any current medication. The experimental protocol was approved by the Institutional Review Board at Drexel University and all participants gave their informed consent. The mean age of the participants was 24.3 years (standard deviation=5.5 years).

EEG activity was recorded from 12 Ag/AgCl electrodes placed at frontal, central, parietal and occipital locations according to the International 10-20 System (F7, F3, Fz, F4, F8, C3, Cz, C4, P3, Pz, and Oz). All electrodes were referenced to linked mastoid leads. Vertical and horizontal electrooculograms (VEOG and HEOG) were monitored via electrodes placed above and below the left eye, and at the left and right outer canthi, respectively. EEG signals were collected using NuAmp amplifier (Neuroscan Inc., El Paso, TX); all impedances were systematically kept below 10 kΩ and the amplification was set to 50 mV/mm. EEG signals were filtered between 0.15 and 100 Hz and sampled at 500 samples/second.

The hemodynamic activity of the prefrontal cortex was recorded using a continuous-wave fNIRS device first described by Chance et al. [22] and further developed at

Drexel University (Philadelphia, PA). The system consisted of three modules: a flexible headpiece, a control box for hardware management and a computer that runs the data acquisition. The headpiece holds four light sources and 10 photodetectors, with a source-detector separation of 2.5 cm, providing a penetration depth of approximately 1.25 cm. The four light sources were activated in turns: each source shone light with input intensity I_0 and the four photodetectors surrounding the currently active source measured the intensity I of the emerging light. The arrangement of sources and detectors on the headpiece and the configuration for data acquisition yields a total of 16 active optodes, which were designed to image cortical areas that correspond to the dorsal and inferior frontal cortices [16]. Each source emitted light at two different wavelengths in the near-infrared spectrum, namely at 730 and 850 nm, and measures of emerging light intensity were obtained for each optode with a sampling frequency of 2 samples/second.

Participants were seated in a dimly-lit, sound attenuated room and were asked to perform a visual *n-back* task, a task widely used to investigate working memory processes [23]. Stimuli were single consonants presented in a pseudo-random sequence on a computer screen. Stimulus duration was 500 ms, with a 2500-ms interstimulus interval. Four conditions were used to incrementally vary the working memory load from zero to three items. In the 0-back condition, subjects responded to a single prespecified target letter (e.g. ''X'') with their dominant hand (pressing a button to identify the target). In the 1-back condition, the target was defined as any letter identical to the one immediately preceding it (i.e., one trial back). In the 2-back and 3-back conditions, the targets were defined as any letter that was identical to the one presented two or three trials back, respectively. The target probability was about 33% for each condition. This strategy incrementally increased working memory load from the 0-back to the 3-back condition. Seven blocks, each containing the four conditions (0-, 1-, 2- and 3-back), were presented to the subjects. The sequence of the four conditions in the seven blocks was randomized. Each presentation of the n-back conditions was followed by a 15 s rest period.

2.3 Data Processing

Information about the behavioral performance in the task was recorded for all subjects. The percentage of correct responses was calculated separately for the four working memory loads and for the overall test. Out of the total pool of subjects, 3 were randomly selected from the group with an overall performance higher than the median ("high performing" group) and 3 were randomly selected from the group with an overall performance below the median ("low performing" group). Table 1 summarizes the behavioral performance for the overall group of subjects.

fNIRS Recordings. fNIRS data were divided into blocks locked to the repeated presentations of the four working memory conditions. Each block lasted 70 s and a 5 s rest baseline was included. The raw data about light absorption acquired by the fNIRS device were low-pass filtered and were converted to changes in concentration of HbO2 and HHb using the modified Beer-Lambert law [24]. The baseline condition used in the modified Beer-Lambert law was the rest period immediately preceding each block. For each of the seven presentations of the 3-back condition, the mean change in HbO2 concentration was extracted. In particular, the channels of most interest were those monitoring the rostral portion of the superior and middle frontal gyri in the left hemisphere, since they have been previously demonstrated to be significantly activated by the n-back task [18]. Therefore, for each subject, average HbO2 values for each of the seven presentations of the 3-back condition were extracted from channels 4, 5 and 6 and used as features in the subsequent classification. Only the 3-back condition was investigated, based on earlier evidence that the oxygenation values recorded during the 3-back condition are affected by the performance level [18].

EEG Recordings. Independent component analysis was used to minimize ocular artifacts in the EEG recordings [25]. Stimulus-locked event-related potentials (ERPs) were extracted from channels Cz and Pz for target stimuli presented in the 3-back condition. A 150 ms pre-stimulus baseline window and a 700 ms post-stimulus response window were used. All epochs were baseline corrected by subtracting the mean of the baseline window form the full epoch. Epochs containing significant movement or muscle artifacts were discarded. The P300 peak was automatically identified at each of the two channels as the largest positive deflection in the 250-600 ms post-stimulus response. For each subject, the average amplitude values of the P300 peak at Cz and Pz were obtained for each of the seven presentations of the 3-back condition and used as features in the classification.

Table 1. Statistics of the behavioral performance (%) in the n-back task for the total pool of subjects [1]

	Mean	(95% Confidence Interval)	Median
0-back condition	90.07 %	(81.29 – 98.84 %)	
1-back condition	92.06 %	(85.63 – 98.47 %)	
2-back condition	85.33 %	(79.23 – 91.42 %)	
3-back condition	78.07 %	(73.76 – 82.37 %)	
Overall	*86.38 %*	*(80.48 – 92.28 %)*	*89.70 %*

Classification. The classification between "high performing" and "low performing" subjects was performed using five different features: two features were obtained from the EEG recordings (the amplitude of the P300 peak at channels Cz and Pz) and three were obtained from the fNRIS recordings (mean change in HbO2 concentration at channels 4, 5 and 6).

For each subject multiple instances of these features were extracted, one for each presentation of the 3-back condition. Each instance ($\boldsymbol{x}_i$) was associated with a label y_i that stated the group of the subject from which the instance was collected (y_i ="low performing" or y_i= "high performing"). The total number of instances $\boldsymbol{x}_i$ was 38 (7 blocks presented to 5 subjects + 3 blocks presented to 1 subject) and constituted the overall set $\boldsymbol{S}$ of available instances: $\boldsymbol{S} = [\boldsymbol{x}_i, y_i]$.

Four different approaches were evaluated to determine their ability to identify "high performing" or a "low performing" individuals:

[1] For one subject, in the "low performing" group, only three of the seven blocks could be presented.

1. *EEG-based classification*: only the features extracted from EEG recordings were used; the feature vector consisted of two elements: the P300 amplitude at Cz and Pz.
2. *fNIRS-based classification*: only the features extracted from fNIRS recordings were used; the feature vector consisted of three elements: the mean change in HbO2 concentration at channels 4, 5 and 6.
3. *Feature-level fusion*: features extracted from both EEG and fNIRS recordings were used and combined in a single feature vector of five elements.
4. *Decision-level fusion*: the classification was performed separately using the EEG and fNIRS features, whose results were then combined to reach the final decision.

Two types of classifiers were investigated to be used in the above mentioned approaches: the Mahalanobis discriminant (MD) and the quadratic classifier (QC). The MD is equivalent to the optimum Bayes classifier if the data are normally distributed with identical (although arbitrary) covariance matrices for all classes [26]. In QC, instances are labeled using a Bayesan error minimization approach, under the more general hypothesis that the covariance matrices for all classes can assume any arbitrary value [26].

For training and testing, a modified k-fold (k=5) cross-validation was implemented. In such an approach, the set $\boldsymbol{S}$ is partitioned into k blocks, each one representing the two groups ("low performing" and "high performing") in an approximately balanced way. Each of the k blocks was in turn held out for testing ($\boldsymbol{S}^{(k)}$), while the other k-1 blocks ($\boldsymbol{S}^{(k-1)}$) were used for training using a bagging procedure [27]. In bagging, an ensemble of classifiers is created: in this study the ensemble was comprised of 10 classifiers all sharing the same architecture but trained on different randomly generated subsets ($\boldsymbol{TS}_{(r)}^{(k-1)}$ r=1,2,…,10) of $\boldsymbol{S}^{(k-1)}$. A label $\hat{y}_{(r)}$ is assigned to each instance in the testing set $\boldsymbol{S}^{(k)}$ by each of the 10 classifiers, which are then combined using a majority voting decision rule. In our implementation, this entire process – of generating 10 training subsets and training 10 corresponding classifiers – was repeated 5 times, each time holding out a different subset $\boldsymbol{S}^{(k)}$ for testing. For each of these 5 repetitions, the *accuracy*, defined as the probability of correctly classifying an instance, was calculated.

2.4 Results

Table 2 summarizes the behavioral performance in the four n-back conditions for the 3 subjects in the "high performing" group and for the 3 subjects in the "low performing" group. The difference in behavioral performance between the two groups is evident in the overall percentage of correctly identified stimuli and in each of the three n-back conditions.

The distribution, in the features space, of the instances collected from the two groups of individuals is presented in Fig. 1A and Fig. 1B. These figures show the features extracted respectively from the EEG recordings (P300 amplitude at Pz and Cz) and from the fNIRS recordings (change in HbO2 concentration at channels 4, 5 and 6); in both spaces the two classes are substantially overlapping. A separate analysis confirmed also a dissociation in the HbO2 values during the 3-back condition between the two groups: whereas for the "high performing" group the HbO2 values were increasing with the working memory load (1-back condition: 0.0086 mM; 2-back condition: 0.0096 mM; 3-back condition: 0.0216 mM), for the "low performing" group this relation was lost and the 3-back condition saw a decrease in the HbO2 values (1-back condition: -0.0213 mM; 2-back condition: 0.0263 mM; 3-back condition: -0.0245 mM).

Table 3 presents the accuracy obtained by the different classification strategies using each of the two classifiers (QD and MD) as base classifiers. In general, the results showed an enhancement in classification performance when features from both EEG and fNIRS are used compared to the results obtained when using them separately. In fact, the mean accuracy for the feature-level fusion and decision-level fusion strategies were overall higher than the mean accuracy for the EEG-based and fNIRS-based classifications.

Table 2. Mean behavioral performance (%) statistics in the n-back task for the "low performing" and "high performing" group. Behavioral performance was calculated as the percentage of presented stimuli that the subject correctly identified as targets or non-targets.

	Low performing	High performing
0-back condition	74.95 %	94.87 %
1-back condition	80.83 %	95.22 %
2-back condition	77.87 %	88.38 %
3-back condition	70.56 %	82.03 %
Overall	76.05 %	90.12 %

Table 3. Mean and standard deviation of the accuracy (%) reached by the quadratic classifier (QC) and Mahalanobis discriminant (MD) used in the four possible classification approaches. Accuracy was evaluated across the 5 repetitions of the k-fold cross-validation procedure.

	QC	MD
EEG-based classification	60.31±18.27 %	62.53±16.44 %
fNIRS-based classification	62.53±21.78 %	51.74± 7.41 %
Feature-level fusion	65.39±16.51 %	58.09±18.56 %
Decision-level fusion	71.11±10.67 %	71.11±10.67 %

Additionally, the two fusion approaches (and in particular the decision-level fusion, in agreement with [28]) provided an increase, albeit small, in the generalizing ability of the classifiers, as measured by a decrement in the accuracy standard deviation.

2.5 Discussions

Overall, the classification of instances collected from "high performing" and "low performing" subjects benefited from the used of combined fNIRS and EEG features. We acknowledge that the results of our statistical analyses cannot be considered conclusive at this time due to the limited data of 3 subjects (from each class) that were available to us. A larger pool of subjects, and therefore a higher number of instances available for training and testing, would allow a better estimation of the accuracy

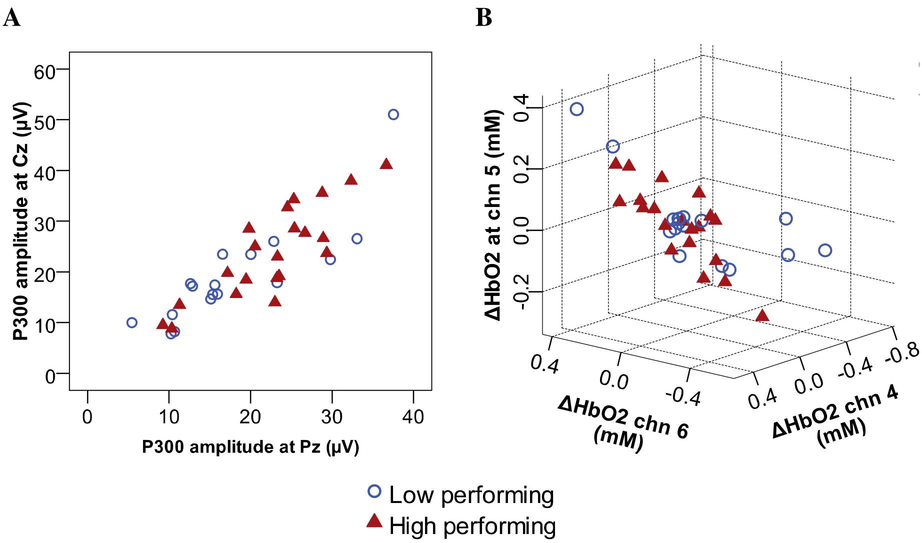


Fig. 1. A. Scatterplot of instances of EEG-based features, the *P300 amplitude at Pz* and the *P300 amplitude at Cz*. B. Scatterplot of instances of fNIRS-based features, the mean changes in HbO2 concentration at channels 4, 5 and 6 (*ΔHbO2 chn4*, *ΔHbO2 chn5* and *ΔHbO2 chn6*, respectively). In both part A and part B, instances are categorized as belonging to *high performing* (*filled triangles*) or *low performing* (*empty circles*) individuals. From each individual, multiple instances were collected, one for each presentation of the 3-back condition.

index. Nonetheless, the combination of fNIRS and EEG improved classification accuracy, even if two relatively simple classifiers where used: a linear parametric classifier (MC) and a nonlinear parametric classifier (QC). It is reasonable to expect that higher accuracies can be obtained using more sophisticated nonlinear classifiers (such as neural networks or support vector machines), that are not bounded by assumptions about the features distributions. Similarly, the relative performance of other fusion algorithms could be investigated, ranging from the simple majority voting (presented in this paper) to multinomial methods, to the fusion of discriminant scores.

3 Concluding Remarks

We have investigated the feasibility and performance of fNIRS and EEG data fusion for the evaluation of the behavioral performance of six healthy adults in a working memory task. Although fNIRS and EEG have been co-registered in previous studies [29], [30], this is the first attempt at their integration by using them together in a pattern recognition application. For this study, the fNIRS-EEG fusion took advantage of both the spatial information about the hemodynamic activity (by including only the channels monitoring the rostral portion of the superior and middle frontal gyri in the left hemisphere) and the fast temporal dynamic of the cognitive processes of interest (by including information about the P300 amplitude). A similar approach can also help explain the mechanisms underlying low task performance in case of neurological disorders such as traumatic brain injury or multiple sclerosis, hence providing some physiological evidence important for the choice of a proper neurorehabilitation and pharmacological intervention. fNIRS-EEG fusion may further be applied to the study of other cognitive domains, in particular taking advantage of the flexibility in task designs allowed by fNIRS and EEG.

References

1. Sergent, J.: Brain-imaging studies of cognitive functions. Trends Neurosci. 17, 221–227 (1994)
2. Stufflebeam, S.M., Rosen, B.R.: Mapping cognitive function. Neuroimaging Clin N Am. 17, 469–484, viii-ix (2007)
3. Malhi, G.S., Lagopoulos, J.: Making sense of neuroimaging in psychiatry. Acta Psychiatr Scand 117, 100–117 (2008)
4. Eliassen, J.C., Boespflug, E.L., Lamy, M., Allendorfer, J., Chu, W.J., Szaflarski, J.P.: Brain-mapping techniques for evaluating poststroke recovery and rehabilitation: a review. Top Stroke Rehabil 15, 427–450 (2008)
5. Calvert, G.A.: Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb Cortex 11, 1110–1123 (2001)
6. Lalanne, C., Lorenceau, J.: Crossmodal integration for perception and action. J. Physiol. Paris 98, 265–279 (2004)
7. Lewine, J.D., Orrison, W.W.J.: Clinical electroencephalography and event-related potentials. In: Orrison, W.W.J., Lewine, J.D., Sanders, J.A., Hartshorne, M.F. (eds.) Functional brain imaging, pp. 327–368. Mosby-Year Book, Boston (1995)
8. Lewine, J.D., Orrison, W.W.J.: Magnetoenecephalography and magnetic source imaging. In: Orrison, W.W.J., Lewine, J.D., Sanders, J.A., Hartshorne, M.F. (eds.) Functional brain imaging, pp. 369–418. Mosby-Year Book, Boston (1995)
9. Raichle, M.E., Mintun, M.A.: Brain work and brain imaging. Ann. Rev. Neurosci. 29, 449–476 (2006)
10. Logothetis, N.K., Wandell, B.A.: Interpreting the BOLD Signal. Annu. Rev. Physiol. 66, 735–769 (2004)
11. George, J.S., Aine, C.J., Mosher, J.C., Schmidt, D.M., Ranken, D.M., Schlitt, H.A., Wood, C.C., Lewine, J.D., Sanders, J.A., Belliveau, J.W.: Mapping function in the human brain with magnetoencephalography, anatomical magnetic resonance imaging, and functional magnetic resonance imaging. J. Clin. Neurophysiol. 12, 406–431 (1995)
12. He, B., Lian, J.: High-resolution spatio-temporal functional neuroimaging of brain activity. Crit. Rev. Biomed. Eng. 30, 283–306 (2002)
13. Ritter, P., Villringer, A.: Simultaneous EEG-fMRI. Neurosci. Biobehav. Rev. 30, 823–838 (2006)
14. Schiff, N.D.: Multimodal neuroimaging approaches to disorders of consciousness. J. Head Trauma Rehabil. 21, 388–397 (2006)
15. Villringer, A.: Functional neuroimaging. Optical approaches Adv. Exp. Med. Biol. 413, 1–18 (1997)
16. Izzetoglu, M., Izzetoglu, K., Bunce, S., Ayaz, H., Devaraj, A., Onaral, B., Pourrezaei, K.: Functional near-infrared neuroimaging. IEEE Trans Neural Syst. Rehabil. Eng. 13, 153–159 (2005)
17. Jobsis, F.F.: Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. Science 198, 1264–1267 (1977)

18. Izzetoglu, M., Bunce, S.C., Izzetoglu, K., Onaral, B., Pourrezaei, K.: Functional brain imaging using near-infrared technology. IEEE Eng. Med. Biol. Mag. 26, 38–46 (2007)
19. Kok, A.: Event-related-potential (ERP) reflections of mental resources: a review and synthesis. Biol. Psychol. 45, 19–56 (1997)
20. Johnson Jr., R.: On the neural generators of the P300 component of the event-related potential. Psychophysiology 30, 90–97 (1993)
21. Kok, A.: On the utility of P3 amplitude as a measure of processing capacity. Psychophysiology 38, 557–577 (2001)
22. Chance, B., Anday, E., Nioka, S., Zhou, S., Hong, L., Worden, K., Li, C., Murray, T., Ovetsky, Y., Pidikiti, D., Thomas, R.: A novel method for fast imaging of brain function, non-invasively, with light. Opt. Express 2, 411–423 (1998)
23. Owen, A.M., McMillan, K.M., Laird, A.R., Bullmore, E.: N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. Hum. Brain Mapp. 25, 46–59 (2005)
24. Delpy, D.T., Cope, M., van der Zee, P., Arridge, S., Wray, S., Wyatt, J.: Estimation of optical pathlength through tissue from direct time of flight measurement. Phys. Med. Biol. 33, 1433–1442 (1988)
25. Jung, T.P., Makeig, S., Humphries, C., Lee, T.W., McKeown, M.J., Iragui, V., Sejnowski, T.J.: Removing electroencephalographic artifacts by blind source separation. Psychophysiology 37, 163–178 (2000)
26. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern classification. Wiley Interscience, New York (2001)
27. Kuncheva, L.I.: Combining pattern classifiers: Methods and algorithms. Wiley Interscience, New York (2004)
28. Parikh, D., Polikar, R.: An Ensemble-Based Incremental Learning Approach to Data Fusion. IEEE Trans Sys. Man Cyber 37, 437–450 (2007)
29. Kennan, R.P., Horovitz, S.G., Maki, A., Yamashita, Y., Koizumi, H., Gore, J.C.: Simultaneous recording of event-related auditory oddball response using transcranial near infrared optical topography and surface EEG. Neuroimage 16, 587–592 (2002)
30. Butti, M., Pastori, A., Merzagora, A., Zucca, C., Bianchi, A., Reni, G., Cerutti, S.: Multimodal analysis of a sustained attention protocol: Continuous Performance Test assessed with Near Infrared Spectroscopy and EEG. In: 28th Annual International Conference of the IEEE EMBS, pp. 1040–1043 (2006)

Transcranial Doppler: A Tool for Augmented Cognition in Virtual Environments

Beatriz Rey[1], Mariano Alcañiz[1], Valery Naranjo[1],
Jose Tembl[2], and Vera Parkhutik[2]

[1] LabHuman, Universidad Politécnica de Valencia, Camino Vera s/n, 46022 Valencia, Spain
{brey,malcaniz,vnaranjo}@labhuman.i3bh.es
[2] Neurology Service, Hospital Universitari La Fe, Av. Campanar 21, 46009 Valencia, Spain
tembl_jos@gva.es, parkhutic_ver@yahoo.com

Abstract. In this work, we propose the use of Transcranial Doppler Monitoring (TCD) as a tool to measure brain activity during the exposure to virtual environments (VE) that can be used in Augmented Cognition (AugCog) systems. The technique is non-invasive, and can be easily integrated with virtual reality (VR) settings. Its high temporal resolution allows the correlation of changes in brain activity to specific events in the VE. In this paper, the TCD technique is described, and results from two studies developed in our group combining TCD with VR are summarized. Possible applications of TCD in the AugCog field are finally discussed.

Keywords: Augmented Cognition, Virtual Reality, Transcranial Doppler, Neurophysiological Data, Cognitive State Assessment.

1 Introduction

Augmented Cognition (AugCog) can be described as "a field of research that seeks to extend a computer user's abilities via technologies that address information-processing bottelenecks inherent in human-computer interaction" [1]. There are two main aspects that should be controlled in an AugCog system. Firstly, the system should be able to detect the instantaneous cognitive state of the user, in order to find limitations such as overload, cognitive lockup, and underload [2]. Secondly, depending on the user's detected state, the system has to adapt the computational interface in order to improve the user's performance.

One of the technologies that can be used in the AugCog field is Virtual Reality (VR). This technology provides users with a controlled environment where different abilities and exercises can be trained. In this context, it is closely related to AugCog and can be seen as a test bed for the technology until it has achieved a sufficient technology readiness level [3].

Previous research in VR has developed environments that can modify dynamically their aspect depending on user's instantaneous state, such as the EMMA system [4]. The EMMA system was designed to be used inside clinical therapy sessions, and has been applied for the treatment of several psychological disorders such as post

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 427–436, 2009.

traumatic stress disorder, pathological bereavement, and adjustment disorder. The main technical feature of this environment is that its aspect can be modified controlled by the therapist that conducts the clinical sessions depending on the emotions that the patient is feeling at each moment, and depending on the purpose of the clinical session. This kind of virtual environment (VE) that can adapt itself to the requirements of the user, instead of having the users to adapt to the device, has been described in the literature as an adaptive display. Adaptive displays have in common with AugCog systems that they have to control the user's instantaneous state and adapt their contents and characteristics depending on this state.

The logical evolution in this kind of systems is that the user's cognitive state can be predicted automatically by the system. Logically, two elements are required for this: sensors for monitoring different physiological signals from users, and an inference engine or classifier to evaluate the information coming from these signals to determine the instantaneous cognitive state of the user. In fact, this is one of the key AugCog research areas: Cognitive State Assessment (CSA) [5]. These sensors acquire "physiological and behavioral parameter(s) that can be reliably associated with specific cognitive states, which can be measured in real-time while an individual or team of individuals is engaged with a system" [5].

Different measurements have been proposed to be related with specific cognitive states of the users [6]. Some of them are based on physiological measurements. For example, electrocardiography [7] and skin conductance [8] have been used to provide estimates of arousal and general cognitive workload. Other behavioral analyses, such as gesture recognition, have been used to monitor levels of attention [9]. Speech and facial expression have also been studied as ways to obtain information about the cognitive state of users [10-11]. However, the most direct analysis of the cognitive state can be obtained using neurophysiological measures. One of the most common measurements in this field is electroencephalogram (EEG), which has been used to estimate various types of cognitive states, such as vigilance, arousal, workload, engagement, distraction and working memory [12]. Other techniques such as functional near infrared spectroscopy (fNIRS) have also been used as a way to measure brain

Fig. 1. Subject with TCD probes placed in their correct location with the help of a probe holder

blood oxygenation and volume changes, helping to understand the cognitive and emotional state of the user during mentally demanding operations [13].

In this paper, we propose another brain activity measurement technique that can be used as a cognitive state sensor: Transcranial Doppler (TCD). It was first used in 1982 as an ultrasound diagnosis technique [14] to control the hemodynamic characteristics of major cerebral arteries in normal and pathological conditions with high temporal resolution. In order to take the measurements, two probes are placed on the skull of the subject using a headband or a probe holder, as can be seen in Fig. 1. Similarly to other cognitive state sensors such as EEG and fNIRS, TCD is a non invasive measurement system in real time, which makes it potentially available for everyday use.

2 Trasncranial Doppler (TCD)

TCD is based on the Doppler Effect. The probes emit an ultrasound beam that is reflected by the blood cells that are moving through the vessels in the brain. The reflected signal is received by the probes with a frequency shift that is proportional to the velocities of the blood cells. TCD studies use mainly the transtemporal window to place the probes that register the ultrasound signal [15]. This window allows the direct registration of the information about the Middle Cerebral Arteries (MCAs), Anterior Cerebral Arteries (ACAs) and Posterior Cerebral Arteries (PCAs). The probe direction, the reference volume depth and the flow direction identify each cerebral artery.

From the reflected signal that is received in the TCD apparatus, it is possible to calculate the maximum blood flow velocity (BFV) at each moment in the arteries under study. Maximum BFV represents the instantaneous velocity of the quickest blood cells. It varies during each cardiac cycle, reaching a maximum shortly after the cardiac contraction and falling to a minimum just before the next. Most TCD studies do not use directly this maximum BFV. Instead, they use an averaged version of it (the mean BFV), which is calculated from the maximum BFV value as the mean velocity value during one complete cardiac cycle [16]. These velocity variations constitute a reliable source of information about brain activity. If the neurovascular coupling is adequate [17], these variations reflect changes in cerebral blood flow (CBF). And previous research has clearly shown that regional CBF increases during the performance of mental activities [18, 19].

TCD has important advantages when compared to other techniques. First of all, it has a high temporal resolution, which allows instantaneous monitoring of cerebral responses to specific events. Furthermore, it is non invasive, so it is possible to use it in an ecological way in a great variety of environments. That constitutes its main advantage when compared with other techniques such as fMRI, which imposes serious restrictions to the experiments in which it is used, as long as the subject has to remain in supine position inside the magnetic resonance machine with minimum head movement while hearing annoying noises.

The main disadvantage of TCD is its spatial resolution, which is limited by the size of the cortical areas supplied by the arteries under study.

2.1 Cortical Areas Supplied by TCD Monitored Arteries

The arteries that supply blood to the greater part of the brain are MCAs. Each MCA carries 80% of the blood flow within its cerebral hemisphere [20]. Their perfusion territory includes subcortical areas, large fractions of the frontal and parietal lobes, as well as the temporal lobes [21], so modifications in MCAs BFV can be produced by different kinds of brain activity (related to motor tasks, attention tasks or emotional states). For example, parts of the parietal and frontal lobe are involved in the processing of emotions, as well as areas of the temporal lobe with the limbic system, so that would justify an increment in MCAs BFV when somebody is experiencing an emotion [22]. In any case, MCAs can provide interesting information about BFV differences between hemispheres in different tasks.

Nevertheless, other vessels that can be monitored with TCD can provide more specific information about the areas that are active at each moment, as long as their perfusion territory is smaller. ACAs supply most of the medial areas of the brain, including the medial frontal cortex and most parts of the limbic system particularly the cingulated [20], so their BFV variations are closely related with the emotional state of the subject who is being monitored. On the other hand, PCAs are the vessels responsible for the irrigation of the primary visual cortex as well as the lateral geniculate body and some of the visual association regions in the occipital cortex [23], so they are useful to analyze brain activity related with visual stimuli.

2.2 TCD in Psychophysiological Studies

TCD has been widely used to monitor cerebral hemodynamics during the performance of cognitive tasks in psychophysiological research. These studies have shown that mean BFV obtained from TCD signals increases when users are doing a cognitive activity when compared to baseline periods [15, 24]. Changes in mean BFV or maximum BFV that can be observed between a baseline period and a phase in which a cognitive task is performed can be analyzed as an absolute change (in cm/s) or as a relative change (in percentage change from baseline). Different cognitive tasks have been considered in these studies, such as reading, arithmetic operations, visual stimulation, attention, verbal tasks, motor tasks, visuospatial tasks and memory.

Some of these studies have analyzed emotion-related changes in BFV. Monitoring emotions can be the basis to obtain a real adaptive display that modifies its aspect depending on the user's emotional state. These studies have found an emotion-related cerebral asymmetry [25, 26], observing a significantly higher increase in the right than in the left MCA during emotional processing.

2.3 TCD in Vigilance Tasks

TCD has also been used in studies closely related with the AugCog field. Several experiences have been related to the study of brain activity during vigilance tasks [27, 28]. Some studies [29, 30] have found that the vigilance decrement in detection rate over time was accompanied by a decrease in BFV in both MCAs. This reduction only happens when the observers are asked to actively monitor the stimuli, and not when they are asked just to look at the vigilance displays with no task to be performed. However, other studies [31] have focused on abbreviated vigilance tasks and, although there was a significant decline in performance over time, there was no significant change in BFV measures over time. This finding does not coincide with earlier findings from long-duration tasks.

TCD has also been used to monitor the influence of automation cues of varying reliability on vigilance performance in a 40 minute simulated air traffic control task [32]. Performance effects for cueing found in the experiment were closely followed by changes in BFV just in the right MCA in conjunction with low salience signals.

3 Transcranial Doppler in Virtual Reality Experiences

Although there have been many previous studies using TCD in neurophysiological research, the first studies that analyze brain activity during a VR exposure have been developed by our group. We have used TCD as a tool to analyze cognitive states related with presence during the exposure of the subjects to a VE in different immersion and navigation conditions. The complete description of the experience can be found elsewhere [33, 34], but a short description of some features of these studies is also included in this section.

3.1 TCD Apparatus and Procedure

A commercially available 2-MHz pulsed-wave TCD unit (Doppler-Box™ Compumedics Germany GmbH) was used to obtain a bilateral continuous measurement of the Doppler signal. This unit allowed the online calculation of BFV during the experiment. The apparatus was connected to a PC in which DWL® Doppler software (QL software) was installed. This software was used to receive the data from the Doppler Box and save the selected variables on the PC hard disk for off-line analysis. Two dual 2-MHz transducers were connected to the Doppler Box. Probes were attached to the user's head using the probe holder provided with the device, as can be observed in Fig. 1. Both hemispheres were simultaneously monitored through the temporal window using two probes capable of simultaneous explorations at two different depths. In the first study [33], both MCAs and ACAs were monitored. The first gate of each probe was located between 50-55 mm depth in order to register left and right MCA (MCA-L and MCA-R) flow. The second gate was located deeper, between 65-70 mm, to take left and right ACA (ACA-L and ACA-R) flow signals. In the second study, [34] only MCAs were monitored.

3.2 Task Description and Virtual Reality Ssetting

The first study [33] was carried out in a CAVE-like environment with four sides (three walls and the floor). The dimensions of the floor were 2.5 x 2.5 m, and the height of the walls was 2.35 m. The device used to navigate was the Flystick (Advance Realtime Tracking GmbH, Weilheim, Germany), which is a wireless joystick with 8 buttons. An optical tracking system, ARTtrack1 (Advance Realtime Tracking GmbH, Weilheim, Germany) was also used to track the user's head and the Flystick position and orientation. The system used active stereoscopy so liquid crystal shutter glasses, CrystalEyes3 (Real D, StereoGraphics, Beverly Hills, USA) were required

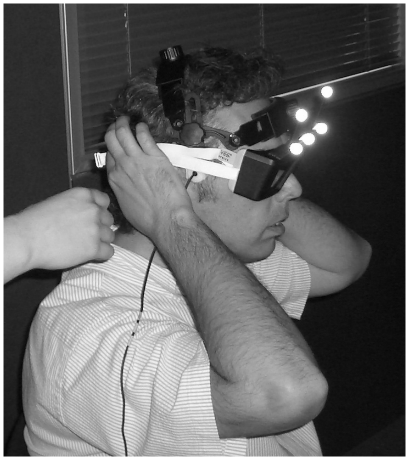


Fig. 2. User with TCD probes and shutter glasses for visualizing VE with stereoscopy

for the visualization. In order to use simultaneously the shutter glasses and the TCD probes, glasses were adapted as can be observed in Fig. 2. No other special modification was required to use TCD in combination with VR hardware.

In the second study [34], some experimental conditions were held in the CAVE-like environment that we have just described. However, in other experimental conditions the environment was retro-projected in a 2 x 1.5 m metacrilate screen. Monoscopic vision was used in this case. The device that was used to navigate inside the VE was an Attack™ 3 Joystick from Logitech (Logitech, Fremont, CA, USA).

In both experiences, a VE composed of several rooms and corridors was used.

3.3 Summary of BFV Variations during the VR Experiences

The first study [33] showed that it was possible to use TCD to monitor brain activity during VR studies. Two different navigation conditions were compared (user-controlled vs. system-controlled navigation). The percentage variations between mean BFV in the user-controlled navigation and its preceding baseline, and between the mean BFV in the system-controlled navigation and its preceding baseline, were positive in all the arteries under study (MCAs and ACAs). The comparison between the percentage variations observed in both navigation conditions showed that significant differences occurred only in the left arteries: MCA-L and ACA-L. The variations in MCA-L could be due to the motor tasks with the right hand to control the Flystick. However, the variations in ACA-L are not directly related to this issue, and can only be explained by other factors such as differences the emotional state or the level of presence that the user is experiencing during the VE exposure in the different navigation conditions. Presence questionnaires confirmed that the level of presence was different between experimental conditions.

The second study [34] compared the same navigation conditions (user-controlled vs. system-controlled) but in two different immersion configurations (corresponding to the two VR settings previously described: CAVE-like vs. projection screen). In this case, only MCAs were considered. The highest BFV percentage variations between VR exposure and its previous baseline were observed in the user-controlled navigation and the CAVE-like configuration. However, only the navigation factor had significant influence in BFV variations. In the case of MCA-L, differences could be mainly due to differences in motor tasks between navigation conditions, as in the previous study. However, differences in MCA-R cannot be explained by these motor tasks, as subjects used the right hand to control the joystick. Exact values of mean MCA-R BFV percentage variations observed in this study are shown in Table 1. A possible explanation of the differences in MCA-R percentage variations between navigation conditions could be found in the different degree of involvement of the users to create a motor plan in both conditions. The level of presence at each condition (which is different as measured by questionnaires) could also be having an influence.

Table 1. MCA-R BFV Percentage Variations (%) in the different experimental conditions in the study about BFV in different navigation and immersion conditions [34]. The mean value and the standard error of the mean (between brackets) are included.

Immersion	Navigation	MCA-R BFV Var. (%)
CAVE-like	User-guided	9.99 (1.28)
	System-guided	8.25 (1.30)
Single screen	User-guided	9.04 (2.00)
	System-guided	3.54 (2.04)

These significant differences between MCA-R percentage variations in the different navigation conditions did not appear in the first study and this is a factor that has to be further analyzed. In any case, it can be observed in Table 1 that differences between navigation conditions are higher in the case of the single screen than in the CAVE-like configuration, and this may be the origin of the significant differences that appear between navigation conditions in the second study. Perhaps in the CAVE-like environment the immersion is so high that brain activity changes that occur are mainly caused by the fact of visualizing the VE inside a CAVE-like configuration, and other aspects such as navigation do not have a big influence in MCA-R BFV. However, in the single-screen condition, the immersion is not so high, and maybe this is why the effect of navigation is not masked and higher differences in percentage variations between navigation conditions are observed.

4 Discussion

These previous studies [33, 34] show that it is possible to use TCD monitoring during the exposure to VE. TCD is a tool that can be easily integrated in VR settings to monitor brain activity during the VR experience. It is possible to obtain reliable TCD signals during the exposure to VE. Besides, the use of TCD does not interfere with the capability of the subjects to focus their attention on the VE. These conclusions show the feasibility of using TCD in combination with VR with different applications such as clinical therapy, tasks training and processes simulation. The main advantage of TCD is that is provides a high temporal resolution that allows the monitoring of fast changes in BFV values caused by neural activity. The use of TCD may help researchers to have reliable information about the brain activity of the subjects and correlate its changes with specific events in the VR session.

The studies also show the feasibility of simultaneously monitoring BFV in different cerebral vessels during the exposure to VE. Two probes are used, so is possible to monitor both hemispheres. In the second study, both MCA-L and MCA-R were studied. However, the selected system also allows the simultaneous monitoring of two different depths (two gates) with the same probe. That allowed us to simultaneously monitor both MCAs and ACAs during the first experience. As different vessels supply different brain areas, an important issue in TCD studies is to identify the brain area most related with the task under study, in order to select the more adequate vessel. It is true that MCAs are the vessels most commonly used because they have the biggest perfusion territory. Studies with MCAs can provide useful information about hemispheric differentiation and temporal evolution, but cannot give detailed information about the specific brain areas that are activated during the task evaluated. However, PCAs and ACAs have smaller perfusion territories, and their spatial resolution can be useful in studies about specific tasks. For example, if visual aspects have to be analyzed, PCAs should be selected because they irrigate the primary visual cortex. If emotional aspects are to be analyzed, ACAs would provide valuable information because they supply most parts of the limbic system and the medial frontal cortex.

Thus, TCD can be used as a cognitive state sensor in AugCog systems. It has already been used to monitor brain activity during vigilance tasks [27 – 32]. Furthermore, it is possible to use it in AugCog systems in which VR setting are used to train different abilities and exercises, combining that way all the different technologies that are discussed in this paper.

Many more applications combining TCD with AugCog systems and VR can be thought, such as memory tasks, attention tasks, motor skills training. VR can provide the virtual laboratory in which all these tasks can be performed and analyzed, and TCD would be the tool to analyze brain activity during task performance to detect changes in cognitive load, underloads and overloads. TCD information would be valuable for the AugCog system to adopt the adequate mitigation strategies and to allocate tasks in an adaptive way.

Closely related with AugCog systems, TCD can be used to advance in the field of adaptive interfaces in VR systems. The EMMA system [4] had the possibility of changing its aspect depending on the emotional state of the patient, but environmental changes were controlled by the therapist that conducted the clinical sessions. However, the use of TCD as a cognitive state sensor combined with this VR system can transform it in an adaptive display that changes depending on the emotional state detected by TCD at each moment. Previous studies have analyzed BFV to monitor emotions [25, 26] and could be the basis of this kind of adaptive display.

Our global conclusion is that TCD is a measure that is worth to study and use as a cognitive state sensor of humans while interacting with computing-based systems.

References

1. Bainbridge, W.S.: Berkshire Encyclopedia of Human-Computer Interaction. Berkshire Publishing Group, Great Barrington (2004)
2. de Greef, T., van Dongen, K., Grootjen, M., Lindenberg, J.: Augmenting Cognition: Reviewing the Symbiotic Relation Between Man and Machine. In: Schmorrow, D.D., Reeves, L.M. (eds.) HCII 2007 and FAC 2007. LNCS (LNAI), vol. 4565, pp. 439–448. Springer, Heidelberg (2007)
3. Vice, J., Lockerd, A., Lathan, C.: Multi-Modal Interfaces for Future Applications of Augmented Cognition. In: Schmorrow, D.D. (ed.) Foundations of Augmented Cognition, pp. 21–27. Lawrence Erlbaum Associates, Mahwah (2005)
4. Alcañiz, M., Botella, C., Rey, B., Baños, R., Lozano, J.A., Lasso de la Vega, N., Castilla, D., Montesa, J., Hospitaler, A.: EMMA: An Adaptive Display for Virtual Therapy. In: Schmorrow, D.D., Reeves, L.M. (eds.) HCII 2007 and FAC 2007. LNCS (LNAI), vol. 4565, pp. 258–265. Springer, Heidelberg (2007)
5. Reeves, L.M., Schmorrow, D.D., Stanney, K.M.: Augmented Cognition and Cognitive State Assessment Technology-Near-Term, Mid-Term, and Long-Term Research Objectives. In: Schmorrow, D.D., Reeves, L.M. (eds.) HCII 2007 and FAC 2007. LNCS (LNAI), vol. 4565, pp. 220–228. Springer, Heidelberg (2007)
6. Grootjen, M., Neerincx, M.A., van Weert, J.C.M.: Task Based Interpretation of Operator State Information for Adaptive Support. In: ACI/HFES 2006, San Francisco (2006)
7. Veltman, J.A., Gaillard, A.W.K.: Physiological workload reactions to increasing levels of task difficulty. Ergonomics 41, 655–669 (1998)
8. Haag, A., Goronzy, S., Schaich, P., Williams, J.: Emotion Recognition Using Bio-Sensors: First Steps Towards an Automatic System. In: André, E., Dybkjær, L., Minker, W., Heisterkamp, P. (eds.) ADS 2004. LNCS (LNAI), vol. 3068, pp. 36–48. Springer, Heidelberg (2004)
9. Ehlert, P.: Intelligent User Interfaces. DKS03-01 Report, Technical University Delft (2003)
10. Rothkrantz, L.J.M., Wiggers, P., van Wees, J., van Vark, R.J.: Voice stress analysis. In: Sojka, P., Kopeček, I., Pala, K. (eds.) TSD 2004. LNCS (LNAI), vol. 3206, pp. 449–456. Springer, Heidelberg (2004)
11. van Kuilenburg, H., Wiering, M., den Uyl, M.J.: A Model Based Method for Automatic Facial Expression Recognition. In: Gama, J., Camacho, R., Brazdil, P.B., Jorge, A.M., Torgo, L. (eds.) ECML 2005. LNCS, vol. 3720, pp. 194–205. Springer, Heidelberg (2005)
12. Freeman, F.G., Mikulka, P.J., Prinzel, L.J., Scerbo, M.W.: Evaluation of an adaptive automation system using three EEG indices with a visual tracking task. Biol. Psychol. 50, 61–76 (1999)
13. Izzetoglu, K., Bunce, S., Onaral, B., Pourrezei, K., Chance, B.: Functional Optical Brain Imaging Using Near-Infrared During Cognitive Tasks. Int. J. Hum.-Computer Interact. 17, 211–227 (2004)
14. Aaslid, R., Markwalder, T.M., Nornes, H.: Noninvasive Transcranial Doppler Ultrasound Recording of Flow Velocity in Basal Cerebral Arteries. J. Neurosurg. 57, 769–774 (1982)
15. Stroobant, N., Vingerhoets, G.: Transcranial Doppler Ultrasonography Monitoring of Cerebral Hemodynamics during Performance of Cognitive Tasks: A Review. Neuropsychol. Rev. 10, 213–231 (2000)
16. McCartney, J.P., Thomas-Lukes, K.M., Gomez, C.R.: Handbook of Transcranial Doppler. Springer, New York (1997)

17. Iadecola, C.: Regulation of the Cerebral Microcirculation during Neural Activity: Is Nitric Oxide the Missing Link? Trends Neurosci. 16, 206–214 (1993)
18. Risberg, J.: Regional cerebral blood flow in neuropsychology. Neuropsychologia 24, 135–140 (1986)
19. Daffertshofer, M.: Functional Doppler testing. In: Hennerici, M., Meairs, S. (eds.) Cerebrovascular Ultrasound, pp. 341–359. Cambridge University Press, Cambridge (2001)
20. Toole, J.F.: Cerebrovascular Disorders, 5th edn. Lippincott Williams & Wilkins, New York (1999)
21. Angevine, J.B., Cotman, C.W.: Principles of neuroanatomy. University Press, New York (1981)
22. Tatu, L., Moulin, T., Bogousslavsky, J., Duvernoy, H.: Arterial territories of the human brain. Cerebral hemispheres. Neurology 50, 1699–1708 (1998)
23. Aaslid, R.: Visually evoked dynamic blood flow response of the human cerebral circulation. Stroke 18, 771–775 (1987)
24. Duschek, S., Schandry, R.: Functional Transcranial Doppler sonography as a Tool in Psychophysiological Research. Psychophysiology 40, 436–454 (2003)
25. Troisi, E., Silvestrini, M., Matteis, M., Monaldo, B.C., Vernieri, F., Caltagirone, C.: Emotion-related cerebral asymmetry: hemodynamics measured by functional ultrasound. J. Neurol. 246, 1172–1176 (1999)
26. Stegagno, L., Patritti, D., Duschek, S., Herbert, B., Schandy, R.: Cerebral blood flow in essential hypotension during emotional activation. Psychophysiol. 44, 232–236 (2007)
27. Warm, J.S., Matthews, G., Tripp, L., Hancock, P.A.: Cerebral Hemodynamics and Brain Systems in Vigilance. In: Schmorrow, D.D. (ed.) Foundations of Augmented Cognition, pp. 707–708. Lawrence Erlbaum Associates, Mahwah (2005)
28. Warm, J.S., Parasuraman, R.: Cerebral Hemodynamics and Vigilance. In: Parasuraman, R., Rizzo, M. (eds.) Neuroergonomics. The Brain at Work, pp. 146–158. Oxford University Press, New York (2007)
29. Mayleben, D.W.: Cerebral Blood Flow Velocity during Sustained Attention. Unpublished doctoral dissertation, University of Cincinnati, OH (1998)
30. Schnittger, C., Sönke, J., Anouschen, A., Münte, T.F.: Relation of Cerebral Blood Flow Velocity and Level of Vigilance in Humans. Cognitive Neurosci. and Neuropsychol. 8, 1637–1639 (1997)
31. Helton, W.S., Hollander, T.D., Warm, J.S., Tripp, L.D., Parsons, K., Matthews, G., Dember, W.N., Parasuraman, R., Hancock, P.A.: The abbreviated vigilance task and cerebral hemodynamics. J. Clin. and Exp. Neuropsychol. 29, 545–552 (2007)
32. Hitchcock, E.M., Warm, J.S., Matthews, G., Dember, W.N., Shear, P.K., Tripp, L.D., Mayleben, D.W., Parasuraman, R.: Automation Cueing Modulates Cerebral Blood Flow and Vigilance in a Simulated Air Traffic Control Task. Theor. Issues in Ergon. Sci. 4, 89–112 (2003)
33. Alcañiz, M., Rey, B., Tembl, J., Parkhutik, V.: A neuroscience approach to virtual reality experience using Transcranial Doppler monitoring. Presence, Teleoperat. & Virtual Environ (in press, 2009)
34. Rey, B., Alcañiz, M., Tembl, J., Parkhutik, V.: Brain Activity and Presence: a Preliminary Study in Different Immersive Conditions Using Transcranial Doppler Monitoring. In: Presence 2008, Padova (2008)

Predicting Intended Movement Direction Using EEG from Human Posterior Parietal Cortex

Yijun Wang and Scott Makeig

Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California, San Diego, USA
{yijun,scott}@sccn.ucsd.edu

Abstract. The posterior parietal cortex (PPC) plays an important role in motor planning and execution. Here, we investigated whether noninvasive electroencephalographic (EEG) signals recorded from the human PPC can be used to decode intended movement direction. To this end, we recorded whole-head EEG with a delayed saccade-or-reach task and found direction-related modulation of event-related potentials (ERPs) in the PPC. Using parietal EEG components extracted by independent component analysis (ICA), we obtained an average accuracy of 80.25% on four subjects in binary single-trial EEG classification (*left* versus *right*). These results show that in the PPC, neuronal activity associated with different movement directions can be distinguished using EEG recording and might, thus, be used to drive a noninvasive brain-machine interface (BMI).

Keywords: posterior parietal cortex (PPC); electroencephalography (EEG); independent component analysis (ICA); brain-machine interface (BMI).

1 Introduction

In current brain-machine interface (BMI) research, predicting intended movement trajectory is a widely proposed method for controlling prosthetic limbs [1]. Most tested systems for monkey and human subjects are based on neuronal activities recorded in the primary motor cortex (M1), where neuron firing patterns encode direction information about limb movement [2-4]. In neuroscience, it is also well known that the parietal cortex plays an important role in movement planning, being involved in sensorimotor transformations from visual input to motor execution. For instance, the posterior parietal cortex (PPC) is critically involved in visuo-motor control of visually guided reaching movements, continuously updating reaching movements to the visual target. According to its role in motor planning, the parietal cortex may provide another way to decode intended movement direction, which can be potential for BMI applications. In recent monkey studies, direction decoding of eye and hand movements has been realized using neuronal signals in the PPC [5]. The PPC of monkey brain can be further divided into subareas for different action planning, e.g., the lateral intraparietal area (LIP) for saccades and the parietal reach region (PRR) for reaches.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 437–446, 2009.

For real-world application, non-invasive brain-computer interfaces (BCIs) based on electroencephalographic (EEG) signals are more practical than invasive BMIs, whose human applications are seriously limited by questions about the safety and durability of implanted electrodes [6-8]. Various EEG signals have been employed to build different kinds of EEG-based BCI systems, e.g., P300 evoked potential, visual evoked potential (VEP), and mu/beta rhythm power [7]. So far, movement direction decoding using noninvasive methods has been tried only in very few studies [9] [10]. In [9], a machine learning paradigm was successfully applied to discriminate movement directions using single-trial EEG data recorded during natural and delayed reaching tasks. However, the functional brain components contributing most to classification have not been specified in this study, and therefore the underlying brain dynamics related to direction coding are still unclear. Recently, a magnetoencephalography (MEG) study showed that the direction of hand movements can be inferred from brain activities [10]. In their study, movement directions were decoded based on power modulation in the low-frequency band (<7Hz) using MEG activities from the motor area. To the best of our knowledge, intended direction decoding in the PPC based on EEG recordings has been rarely studied, and is ignored in current BCI research. In the present study, we investigated brain activity in the human PPC during directional movement planning using multichannel event-related potentials (ERPs), and propose a BCI scheme based on single-trial EEG classification.

2 Method

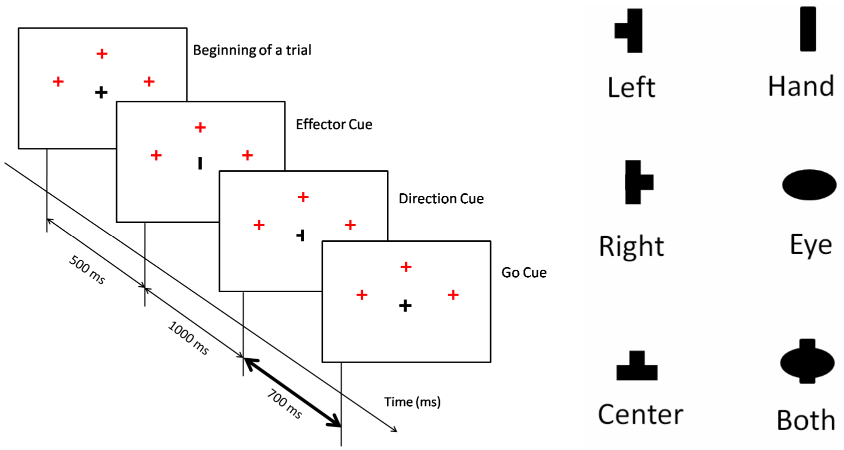
2.1 Subjects

Four healthy, right-handed participants (3 males and 1female, mean age 25 years) with normal or corrected-to-normal vision performed this experiment. All participants were asked to read and sign an informed consent form approved by the UCSD Human Research Protections Program before participating in the study.

2.2 Stimuli and Procedure

During execution of eye or hand movements, movement artifacts including electrooculographic (EOG) and electromyographic (EMG) signals also include direction information about the attended movement. To obtain clean brain signals not including such information, therefore, a delayed saccade-or-reach task was used in this study, allowing us to look for direction information in the EEG during the phase of movement planning. The experiment was comprised of nine conditions differing by movement type (saccade to target, reach without eye movement, or visually guided reach) and movement direction (left, center, or right). Each task was indicated to the subject by, first, giving an effector cue telling the type of action to be performed, followed by a direction cue and, finally, by an imperative action cue. Subjects were seated comfortably in an armchair at a distance of 40cm from a 19-inch touch screen. A chin rest was used to help them maintain head position.

Subjects used their right hand to perform reach tasks. At the beginning of each trial, the subject's forearm rested on the table with index finger holding down a key on a keypad placed 30cm in front of screen center. The sequence of visual cues in each trial is shown in Fig.1(a). At the beginning of a trial, a fixation cross (0.65°×0.65°) was displayed in the center of the screen plus three red crosses (0.65°×0.65°) indicating potential target positions. The left and right targets had a vertical distance of 6° and a horizontal distance of 15° from the central fixation cross; the central target was 12° upwards. After 500ms, an effector cue (0.5°×0.5°, rectangle, ellipse indicating hand and eye movements respectively, see Fig.1(b)) appeared at screen center for 1000ms. Next, a central direction cue (0.65°×0.65°, ┤, ┴, ├ for left, center, and right respectively) was presented for 700ms. Subjects were asked to maintain fixation on the central cue until they started their response, to perform the indicated response as quickly as possible following the disappearance of the direction cue (and reappearance of the fixation cross), and finally to return to their initial (key-down) position. Total trial duration amounted to 3500~4000ms.



(a) (b)

Fig. 1. (a) Time sequence of cue presentation in a trial and (b) visual cues used to indicate effector and direction of a task. In each trial, three central cues (first, effector cue, next, direction cue, and finally, go cue) were presented. The 700 ms delay period between the "Direction cue" and "Go cue" was considered the phase of directional movement planning. EEG data segment within this period was used for further analysis.

Auditory feedback was given to help the subjects fulfill the instructions correctly. Four different tones were used to mean "correct", "error", "early", and "time out", respectively. In the reach tasks, if the point on the screen touched by the subject was outside the boundary of a (5.5°×5.5°) square centered on the target cross, the "error" tone sounded. If the response began during the movement preparation period (0-700ms after direction cue onset), the "early" warning sounded. The "time out" feedback sounded when response time was >500ms. All other trials were followed by the "correct" feedback sound. Only those trials are considered here. Subjects were instructed to perform tasks accurately to achieve a high score (percentage of correct trials). Their score was displayed on the screen at the end of each block. Some practice blocks were run before starting the EEG recording. For each subject, the experiment consisted of four blocks (with breaks in between) each including five runs of 45 trials. Within each block, there was a 20-second rest between runs. A total of 900 trials were equally distributed between the nine tasks, which were presented to the subject in a pseudorandom sequence.

2.3 Data Recording

EEG data were recorded using Ag/AgCl electrodes from 128 scalp positions distributed over the entire scalp using a BioSemi ActiveTwo EEG system (Biosemi, Inc.). Eye movements were monitored by additional bipolar horizontal and vertical EOG electrodes. All signals were amplified and digitized at a sample rate of 256 Hz. Electrode locations were measured with a 3-D digitizer system (Polhemus, Inc.). Three cue presentation events and two manual response events ("button release" and "screen touch") were recorded on an event channel synchronized to the EEG data by DataRiver software (A. Vankov).

2.4 Data Processing and Analysis

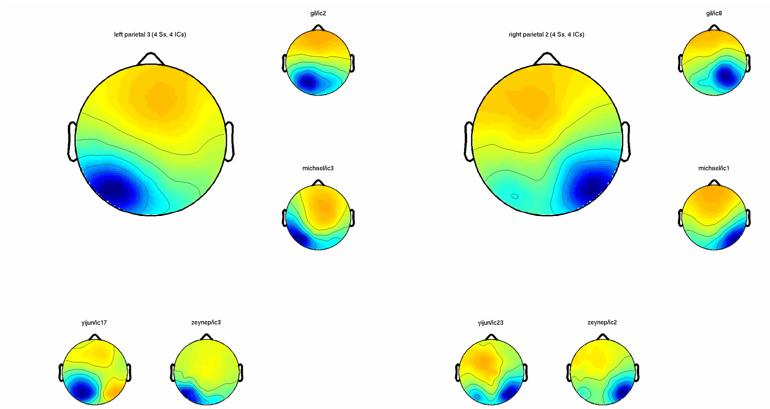
Here, we only focused on estimations of planned direction of movement. Therefore we first separated the trials for each subject into three classes (left, right, and center) for offline analysis. In each class, the three tasks with different effectors (hand, eye, both) were mixed together. Investigation of effector-specific (hand or eye) EEG activations will not be included in this paper.

Data were analyzed using tools in the EEGLAB toolbox [11]. Epochs from the response delay period, 0 to 700ms following direction cue onset, were extracted from the continuous data, and labeled by movement direction. The period [0, 100ms] was used as baseline for each trial. Electrodes with poor skin contact were identified by their abnormal activity patterns and then removed from the data. For each subject, electrode locations were co-registered with a spherical four-shell head model used for dipole source localization.

Spatial Filtering

Independent component analysis (ICA) has been widely used in EEG analysis [12-14]. It can decompose the overlapping source activities constituting the scalp EEG into functionally specific component processes. Here, we used ICA as an unsupervised spatial filtering technique to extract parietal EEG independent component (IC) activities that excluded noise from eye and muscle components as well as brain activities from other functional processes (e.g., in motor, visual, and frontal areas). For each subject, all trials were band-pass filtered (1-30 Hz), concatenated, and then decomposed using the extended infomax ICA algorithm [15]. Two lateralized temporo-parietal components were easily identified in each subject's decomposition by their spatial projections and significant contributions to the average event-related potential (ERP) waveforms time locked to onsets of the movement direction cue.

Figure 2 shows the scalp projections of the two parietal component clusters for all four subjects, plus their mean scalp maps. Clustering was done based on IC scalp maps using EEGLAB tools. These components contributed most to the scalp ERPs obtained by averaging the channel data over all the trials. To indicate the anatomical source location of these components, IC maps were subjected to equivalent dipole localization using the EEGLAB plug-in DIPFIT [11]. Source locations were specified in the Talairach coordinate system. Equivalent dipole localization (average Talairach coordinates: [-33, -59, 28] in the left hemisphere and [40, -49, 30] in the right hemisphere) indicated that these IC sources originated from the PPC (Brodmann Area 39/40). These results demonstrate that the PPC is activated during intended movement planning. To further explore the underlying neural mechanism of direction coding in the PPC, the parietal ICs were selected and back-projected onto the scalp to visualize their separate contributions to the scalp data.



(a) (b)

Fig. 2. Two clusters of lateralized temporo-parietal components with equivalent dipole locations in the (a) left hemisphere and (b) right hemisphere. Large cartoon heads show the mean scalp map for each cluster. Small heads show the clustered component maps for each of the four subjects.

ERP Modulation

To extract the direction-specific portion of the ERPs, we compared the spatiotemporal patterns of the parietal EEG components for the different movement directions. For all four subjects, we found a consistent hemispheric asymmetry over the parietal cortex during the delay period (0-700ms, 0-100ms used as baseline) in which motor planning can be presumed to have continued until cued movement onset (after 700ms). The projected PPC ICs produced a significant contralateral negativity and ipsilateral positivity with respect to intended movement direction. Scalp maps of left, right, and center classes for one subject were shown in Fig.3. For the “left” and “right” classes, their maps showed significant ipsilateral positivity. For instance, the left hemisphere has much higher amplitude than the right hemisphere when planning left movements. For the “center” condition, the map has a symmetric distribution on both sides and the amplitudes are much lower compared to “left” and “right” conditions. To further investigate the time course of this hemispheric asymmetry, difference wave was calculated by subtracting the contralateral activity from the ipsilateral activity with respect to movement direction. Two electrodes with highest weights in

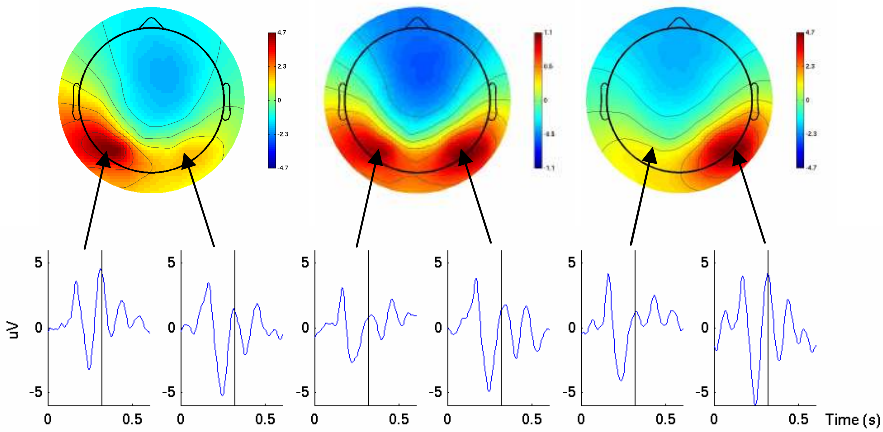


Fig. 3. Scalp maps and ERP waveforms of the summed, back-projected parietal ICs for one subject in the three different direction conditions (left, center, and right) at 320ms after the direction cue. Note that the color scales of the scalp maps differ. The ERP waveforms were from two lateral parietal electrodes with strongest PPC projections.

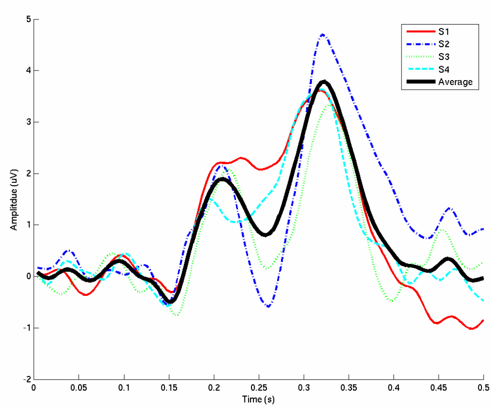


Fig. 4. Ipsilateral minus contralateral difference waves averaged over the "left" and "right" trials. Two peaks centered at 200ms and 320ms were the most significant hemispheric asymmetries appearing during planning of directional movements.

the two parietal IC maps were selected to represent the left and right hemispheres. In the difference wave averaged across the "left" and "right" trials, the hemispheric asymmetry was characterized by two contralateral negativities peaking 200ms and 320ms after the direction cue respectively, with mean amplitudes of 1.9μV and 3.8μV across subjects (see Fig.4).

Feature Extraction and Classification

As a first evaluation of the potential use of EEG activity in PPC for driving a BCI system, binary classification of "left" versus "right" trials was performed using standard machine learning techniques that have been successfully employed in current BCI research [16-18]. Because this study focused on EEG modulation in the parietal cortex, only the parietal IC components were used for feature extraction, although other cortical ICs might contribute separate information for classification of intended direction (e.g., somatomotor components). Although subjects were instructed not to make any response during the movement planning period, covert eye and muscle movements might have occurred, giving additional EEG signals informative for classifying movement direction contained in ICs accounting for eye or scalp muscle activities. Here we constrained the classification performance to reflect only the directional EEG information generated in parietal cortex. Subject-specific time- and frequency-domain parameters were derived for classification. A sliding window was used to optimize the latency and frequency windows giving best classification performance. Because we found that the low-frequency activity contributed to the classification for all subjects, for simplicity a low-pass filter was used to extract the frequency components. The selected time/frequency parameters were listed in Table.1. Not unexpectedly, optimized time windows are consistent with the time course character of the difference wave shown in Fig.4.

After low-pass filtering, normalized amplitudes in the selected time window, normalized at each time point to have a range of [-1 1] across trials, were employed as features. Feature vectors from both parietal components were concatenated and then input to a support vector machine (SVM) classifier using an RBF kernel. The SVM algorithm was performed using the LIBSVM toolbox [19]. 10x10-fold cross validation was run to estimate classification performance.

3 Results

We used classification accuracy to evaluate classification performance. An average accuracy of 80.25±2.22% was obtained for single-trial classification across the four subjects. The classification results are listed in Table 1. Considering that this paradigm is based on single-trial classification, the accuracy is comparable to most current BCI systems, e.g., the P300-based and motor imagery-based BCIs. Moreover, subject variety (reflected in the standard deviation across the four subjects of only 2.22%) does not appear to be as large as in other BCI system reports, suggesting that this method might be usable for more subjects than the other BCI systems. Testing this impression would doubtless require more subjects. These results suggest that more refined measures of movement intention-related EEG activity arising in the PPC (and elsewhere in cortex) might be used to build a robust and noninvasive BCI system.

Table 1. Time-frequency parameters and classificatioin performance for all subjects

Subject	Time Window (ms)	Frequency Window (Hz)	Accuracy (%)
S1	180-500	0-25	79.9±0.45
S2	150-480	0-25	81.9±0.94
S3	180-450	0-20	77.2±0.86
S4	210-510	0-35	81.9±0.81
Mean	——	——	80.25±2.22

4 Conclusion and Discussions

In this EEG study, we designed a movement delay paradigm to investigate brain activities in the human PPC during planning of intended movements. The results indicated that EEG signals generated in the PPC are altered during movement planning, and their hemispheric asymmetries carry information about intended movement direction. By analyzing multi-channel ERPs at the single-trial level, we obtained stable classification of "go left" and "go right" planning trials for all subjects. The resulting classification accuracy of 80.25% makes this paradigm promising for BCI design.

Classification performance might be improved by considering the following factors. First, during motor planning, the PPC also encodes effector information, producing effector-specific brain activity patterns [20]. In the current data analysis, three tasks with different kinds of effectors (hand, eye, and both) were not distinguished, and may introduce variance linked to the different effectors used. Therefore, classifying trials involving the same effector might be more efficient. Else, a multi-factorial classification scheme might be used that included information as to the intended effector. Finally, the same data might be able to predict both the intended effector and movement direction. Second, for feature extraction a simple sliding window was used to select the latency window and frequency band used. To find more informative parameters, time-frequency decomposition methods might be applied allowing additional selection of optimal time-frequency measures. Third, additional features derived from EEG power modulation may be complementary to current features obtained from the time-domain waveforms. For example, [21] showed that the direction of visuospatial attention could be predicted by measuring alpha band power over the two posterior brain hemispheres.

Several potential applications of this paradigm may be expected. It could be directly used to implement a BCI based simply on decoding movement direction. Else, it could be integrated into current BCI systems to realize more robust or multi-dimensional control. For example, combining this paradigm with a motor imagery-based BCI (using EEG changes linked to imagining movements of left hand and right hand), might double the number of selective commands (from 2 to 4). Else, motor imagery of left and right hand movements might be linked to different directions (e.g., by imagining the left hand pointing to the left, or the right hand pointing to the right). In this case, by introducing additional parietal EEG components to mu/beta components from sensorimotor areas, classification performance can be significantly increased, although in this case the system would remain a two-class mode.

Before implementing a practical online BCI system based on intended movement direction, several issues still need further investigation. First, changes in attention and intention both contribute to direction-related EEG modulation. To learn more details about the relationship between these two factors, standard spatial attention experiments might be used to identify purely intention-related features. Else, some combination of subject attention and intention might give more efficient direction-specific brain patterns for a BCI communication or control system. In a practical system, movement planning without subsequent motor activity might be associated with lower BCI performance. The participants in this study were healthy volunteers; a direct test of the system concept on patients with motor disabilities will therefore be necessary before proposing applications for subjects with motor disabilities.

References

1. Lebedev, M.A., Nicolelis, M.A.L.: Brain-Machine Interfaces: Past, Present and Future. Trends in Neurosciences 29(9), 536–546 (2006)
2. Taylor, D.M., Tillery, S.I.H., Schwartz, A.B.: Direct Cortical Control of 3D Neuroprosthetic Devices. Science 296, 1829–1832 (2002)
3. Nicolelis, M.A.L.: Actions from Thoughts. Nature 409, 403–440 (2001)
4. Hochberg, L.R., Serruya, M.D., Friehs, G.M., Mukand, J.A., Saleh, M., Caplan, A.H., Branner, A., Chen, D., Penn, R.D., Donoghue, J.P.: Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature 442(7099), 164–171 (2006)
5. Quiroga, R.Q., Snyder, L.H., Bastista, A.P., Andersen, R.A.: Movement Intention Is Better Predicted than Attention in the Posterior Parietal Cortex. J. Neurosci. 26(13), 3615–3620 (2006)
6. Wolpaw, J.R., Birbaumer, N., Heetderks, W.J., McFarland, D.J., Peckham, P.H., Schalk, G., Donchin, E., Quatrano, L.A., Robinson, C.J., Vaughan, T.M.: Brain-Computer Interface Technology: A Review of the First International Meeting. IEEE Trans. Rehabil. Eng. 8, 164–173 (2000)
7. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain-Computer Interfaces for Communication and Control. Clinical Neurophysiology 113(6), 767–791 (2002)
8. Birbaumer, N.: Breaking the Silence: Brain-Computer Interfaces (BCI) for Communication and Motor Control. Psychophysiology 43, 517–532 (2006)
9. Hammon, P.S., Makeig, S., Poizner, H., Todorov, E., de Sa, V.R.: Predicting Reaching Targets from Human EEG. IEEE Signal Processing Magazine 25(1), 69–77 (2008)
10. Waldert, S., Preissl, H., Demandt, E., Braun, C., Birbaumer, N., Aertsen, A., Mehring, C.: Hand Movement Direction Decoded from MEG and EEG. J. Neurosci. 28(4), 1000–1008 (2008)
11. Delorme, A., Makeig, S.: EEGLAB: An Open Source Toolbox for Analysis of Single-Trial EEG Dynamics Including Independent Component Analysis. J. Neurosci. Meth. 134, 9–21 (2004)
12. Makeig, S., Westerfield, M., Jung, T.P., Townsend, J., Courchesne, E., Sejnowski, T.J.: Dynamic Brain Sources of Visual Evoked Responses. Science 295, 690–694 (2002)
13. Jung, T.P., Makeig, S., McKeown, M.J., Bell, A.J., Lee, T.W., Sejnowski, T.J.: Imaging Brain Dynamics Using Independent Component Analysis. Proc. IEEE 89, 1107–1122 (2001)
14. James, C.J., Hesse, C.W.: Independent Component Analysis for Biomedical Signals. Physiol. Meas. 26, R15–R39 (2005)
15. Lee, T.W., Girolami, M., Sejnowski, T.J.: Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Subgaussian and Supergaussian Sources. Neural Comput. 11(2), 417–441 (1999)
16. Lotte, F., Congedo, M., Lecuyer, A., Lamarche, F., Arnaldi, B.: A Review of Classification Algorithms for EEG-Based Brain-Computer Interfaces. J. Neural Eng. 4, R1–R13 (2007)
17. Müller, K.R., Krauledat, M., Dornhege, G., Curio, G., Blankertz, B.: Machine Learning Techniques for Brain-Computer Interfaces. Biomed. Tech. 49(1), 11–22 (2004)
18. Kaper, M., Meinicke, P., Grossekathoefer, U., Lingner, T., Ritter, H.: BCI Competition 2003-Data Set IIb: Support Vector Machines for the P300 Speller Paradigm. IEEE Trans. Biomed. Eng. 51(6), 1073–1076 (2004)

19. Chang, C., Lin, C.: LIBSVM : a Library for Support Vector Machines (2001), `http://www.csie.ntu.edu.tw/~cjlin/libsvm`
20. Calton, J.L., Dickinson, A.R., Snyder, L.H.: Non-Spatial, Motor-Specific Activation in Posterior Parietal Cortex. Nat. Neurosci. 5, 580–588 (2002)
21. Thut, G., Nietzel, A., Brandt, S.A., Pascual-Leone, A.: Alpha-Band Electroencephalographic Activity over Occipital Cortex Indexes Visuospatial Attention Bias and Predicts Visual Target Detection. J. Neurosci. 26(37), 9494–9502 (2006)

Enhancing Text-Based Analysis Using Neurophysiological Measures

Adrienne Behneman[1], Natalie Kintz[1], Robin Johnson[1], Chris Berka[1], Kelly Hale[2], Sven Fuchs[2], Par Axelsson[2], and Angela Baskin[2]

[1] Advanced Brain Monitoring, 2237 Faraday Ave., Suite 100, Carlsbad, CA 92008
[2] Design Interactive Inc., 221 E. Broadway, Suite 110, Oviedo, FL 32765
{abehneman,nkintz,rjohnson,chris}@b-alert.com

Abstract. Intelligence analysts are faced with the demanding task of identifying patterns in large volumes of complex, textual sources and predicting possible outcomes based on perceived patterns. To address this need, the Advanced Neurophysiology for Intelligence Text Analysis (ANITA) system is being developed to provide a real-time analysis system using EEG to monitor analysts' processing of textual data during evidence gathering. Both conscious and unconscious 'interest' are identified by the neurophysiological sensors based on the analyst's mental model, as related to specific sentences, indicating relevance to the analysis goal. By monitoring the evidence gathering process through neurophysiological sensors and implementation of real-time strategies, more accurate and efficient extraction of evidence may be achieved. This paper outlines an experiment that focused on identifying distinct changes in EEG signals that can be used to decipher sentences of relevance versus those of irrelevance to a given proposition.

Keywords: EEG, Reading, Relevancy, Alpha, Theta.

1 Introduction

Textual data mining for intelligence analysts involves deriving high quality information based on relevance, novelty, and distinctiveness from large volumes of complex textual sources. The challenge is in transforming information from these unstructured and massive collections into small and precise chunks suitable for reasoning [1]. Scanning immense quantities of data can be tedious and takes time away from the goal of using this information to draw a conclusion. In fact, analysts often spend the majority of their time finding the correct information associated with their research question, leaving little time for analyzing and projecting possible outcomes.

Given that analysts are a key component in the analytic processing of text sources, it is important to devise tools that can aid them in both top-down and bottom-up analysis processes. There are a multitude of software systems available and/or being developed, designed to search through text sources and focus an information analyst on the nugget of information that is important. However, as stated by Cowell (2006), "the majority of analysts use the same techniques they learned in graduate school,

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 449–458, 2009.

including printing out hard copies and highlighting, or copying sources into an electronic Word document and arranging material into the required template." [2] This reliance on these basic processes suggests that the tools provided do not target the needs of the analyst. Thus, a distinct point of opportunity is apparent.

An automated intelligent system could support the data foraging stage [3] of the analysis process, within which analysts search vast amounts of information for chunks of evidence that may be buried in various sources. Analyst bias and/or inattention may enter at this early stage and inhibit the extraction of information so that it is unavailable for future hypothesis generation. In order to avoid problems of early evidence rejection, a real-time closed-loop system could be implemented that monitors text processing and associated decision making (i.e., selecting relevant data to include in analysis) and identifies and tracks subconscious 'interest' in unselected text and potential cases for mitigation (e.g., potentially relevant information discarded based on top-down processing or 'explaining away'). Such a tool would ensure that all relevant information (both supporting and opposing the original generated hypotheses) is taken into account when generating hypotheses during later stages of the information analysis process and potentially reduce the effects of analyst bias and inattention.

A real-time analysis system (ANITA) is being developed that uses EEG to monitor analysts' processing of textual data during evidence gathering. The neurophysiological relevance indicator, based on changes in sub-bands of EEG frequencies, is used to auto-extract text snippets from reviewed documents that are relevant to the current analysis goal as reflected in the analyst's mental model, while still allowing the analyst to also manually extract items they perceive as relevant. By monitoring the evidence gathering processes through neurophysiological sensors and implementation of real-time strategies, a more accurate, faster and less biased extraction of evidence may be achieved.

2 Methods

A software-based test-bed presented a series of text analysis scenarios to participants in which single sentences were presented on digital cue cards. The test-bed allowed for real-time data synchronization of EEG and behavioral responses (e.g., key presses). The time boundaries identifying the participants' mental processing of each individual sentence were determined by behavioral responses, and were used to develop distinct, event-specific neurophysiological indicators to indicate the 'relevance' versus 'irrelevance' of assessed information to a provided proposition (analysis question).

2.1 Participants

A total of 27 healthy subjects (15M/ 12F), with an average age of 26.5 (s.d. 8) participated in the experiment. Eighteen participants completed the experiment at the offices of Design Interactive, Inc. in Oviedo, FL, and 9 participants completed the study at the offices of Advanced Brain Monitoring, Inc. in Carlsbad, CA. Participants were free from a history of neurological, psychiatric and attention deficit/hyperactivity disorders, head trauma, use of psychotropic or illicit drugs, or abnormal sleep patterns and sleep quality. No participants were pregnant, nor did they excessively consume alcohol (>5drinks/day) or caffeine, (>800mg/day). The rationale for use of the screening criteria was to exclude conditions that may affect the EEG. All participants had normal or corrected-to-normal vision.

2.2 Apparatus

All participants wore the wireless B-Alert® EEG Sensor Headset developed by Advanced Brain Monitoring (ABM), a portable system to record EEG signals as well as heart rate. The initial 18 participants wore the 6-channel differential EEG configuration with electrodes located at Fz, Cz, POz, F3, C3, and C4, according to the international 10-20 system [4]. From these electrode sites, five differential channels were collected (Fz-POz, Cz-POz, C3-C4, Fz-C3, and F3-Cz). The subsequent nine participants wore the 9-channel referential configuration with electrodes located at Fz, Cz, POz, F3, F4, C3, C4, P3, P4; linked reference electrodes were located behind each ear on the mastoid bone. The EEG signal was sampled at a frequency of 256 Hz. To capture eye gaze and pupil size, a stand-alone non-intrusive Near-Infrared (NIR) eye-tracking system was used. Eye tracking results are outside the scope of this paper, and will be reported elsewhere.

A PC was used to drive visual presentation of experimental conditions on a flat panel monitor. Stimuli from the PC were time-synched to the EEG data collection system. Investigators used a Java-implemented experimental software-based text analysis environment that captures the data stream from the B-Alert EEG-Headset to record user neurophysiological data. The test-bed allows for real-time data synchronization and creation of log files of EEG and behavioral response data (e.g., key presses).

2.3 Experimental Tasks

Participants were first shown a short background story that provided the analysis scenario and a related one-sentence proposition (analysis question), and were then asked to view a series of sentences to determine their relevance to the provided proposition. The scenarios were constructed based on two case studies created by Dr. Frank Hughes, professor in the Department of Intelligence Research and Analysis at the Joint Military Intelligence College (JMIC) [5, 6]. Based on solutions provided by Dr. Hughes, 10 relevant (R) sentences were identified. For each proposition, ten additional sentences were created that were completely irrelevant (CI) to the analysis question. Finally, 10 sentences were added to each analysis question that were not relevant for the analysis but contained some of the key words found in the proposition or case study topic (referred to as semi-irrelevant, SI). Thus, 30 sentences (10 of each relevance level) were associated with each analysis question/proposition. The test-bed presented sentences from a narrative text, one at a time, on digital cue cards (Fig. 1), randomized in terms of relevance level, and allowed the participant to either rate the level of interest (directly *relevant* or *not relevant* to previously provided proposition) of each item or advance to the next sentence if they were not yet ready to make a decision. Using this approach, physiological indicators of relevance were collected at

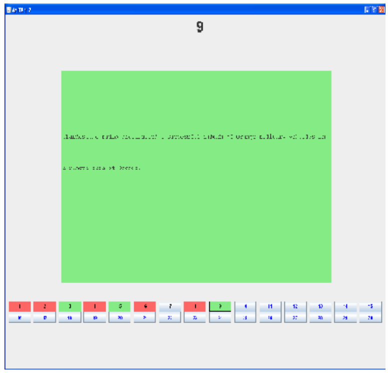


Fig. 1. Sentence-by-Sentence Test-bed Screenshot

the sentence level (i.e., with minimal interference from adjacent textual content), in order to develop template signatures of relevance.

The test-bed presented each sentence as a card on which the participant could click to indicate relevance or irrelevance with regard to the analysis question. Left-clicking a sentence rated it relevant, changing the color of the card and the associated thumbnail at the bottom of the screen to green. Right-clicking a sentence rated it irrelevant, changing the color of the card and the associated thumbnail at the bottom of the screen to red. If a sentence was left unrated, the card and associated thumbnail remained gray, but the number on the thumbnail changed from blue to black to indicate the evidence had been previously viewed. Clicking on previously viewed sentences was possible to change responses if desired, but moving forward was only possible in numerical order (i.e., go to the next unviewed card).

2.4 Analysis Procedure

EEG and test-bed data streams were synchronized in real time. EEG data was analyzed based on participant responses; where clicking on a cue card indicated the beginning of sentence processing, and a relevance decision (participant rating the sentence as either relevant or irrelevant) indicated the end of sentence processing.

To analyze EEG data, identification and decontamination of spikes, amplifier saturation, and environmental artifacts was accomplished using methods described in Berka, 2004 [7]. The EEG signal was then band pass filtered to select for the following frequencies: slow theta (3-5Hz), fast theta (5-7Hz), total theta (3-7Hz), slow alpha (8-10Hz), fast alpha (10-12Hz) and total alpha (8-12Hz). To measure the change in each EEG frequency band related to sentence processing, the rate of change in power of each band was measured for each EEG channel during (a) entire sentence processing, (b) the first second of sentence processing, and (c) the last second of sentence processing (the second preceding a response). The rate of change in power of a given frequency band was calculated by fitting a line through a sequence of data points representing the time evolution of the power of the band over the selected period (whole sentence, the first or last second). The slope of the derived regression line was taken as a measure of the rate of change in power of the analyzed frequency band. A positive linear regression value indicates an increase (synchronization) in a frequency band, while a negative linear regression value reflects a decrease (desynchronization). Only items in which the participant responded correctly the first time a sentence was viewed were included in the current analyses.

3 Results

EEG analysis focused on single trial evaluation of linear trends in theta and alpha activity in order to identify template signatures of processing a relevant item as compared to an irrelevant item. The 6-channel differential configuration (used to collect data from n=18) provides a relatively global view of EEG activity, while the 9-channel referential configuration (n=9) allows for more localized and lateralized analyses.

3.1 Data Collected with 6-Channel Eeg Configuration

The slope of regression lines for each frequency sub-band was calculated for each differential channel during the entire sentence, first second, and last second of processing. An ANOVA revealed a significant effect in slow theta (3-5Hz) for channel CzPOz that distinguishes the slope of the regression line for CI sentences from R and SI , $F\,(2{,}1455) = 4.535$, $p < 0.05$. As seen in Fig. 2, CI sentences had significantly greater synchrony in CzPOz than both the R and SI sentences; however R and SI sentences were not different from each other

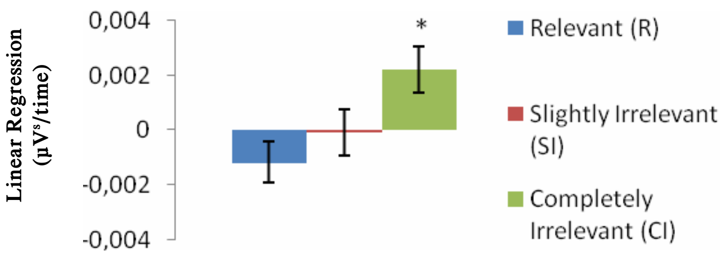


Fig. 2. Slope of regression lines of the Slow Theta band (3-5 Hz) during processing of R, SI, and CI sentences, averaged across 9 participants. Measurement began at time of sentence presentation, and ended at time of keyed response.

CI sentences showed an increase in slow theta power from start of sentence presentation until response. SI sentences showed little change in power, and R sentences showed a negative change in power. Because the differential channels emphasize differences between brain regions, but make changes occurring synchronously in the two regions less visible, the team collected further recordings with the 9-channel monopolar system.

3.2 Data Collected with 9-Channel Eeg Configuration

The 9-channel referential EEG configuration reveals more localized and lateralized patterns. As can be seen in Fig. 3, the fast theta band (5-7 Hz) *increased* in power during the processing of CI sentences more than during the processing of R or SI sentences. The effect was most prominent at the right-hemisphere sites (F4, C4, P4), and at the posterior sites (P3, POz, P4). An ANOVA revealed significant mean differences between fast theta linear regression values of sentences of differing relevancy at electrode sites F4, C4, P4, and POz [F's $(2,370) \geq 3.382$, p's ≤ 0.05]. The effect was strongest at POz, $F\ (2,370) = 5.684$, $p < 0.01$. Although CI was significantly more synchronized compared to both R and SI categories, R and SI were not significantly different from each other.

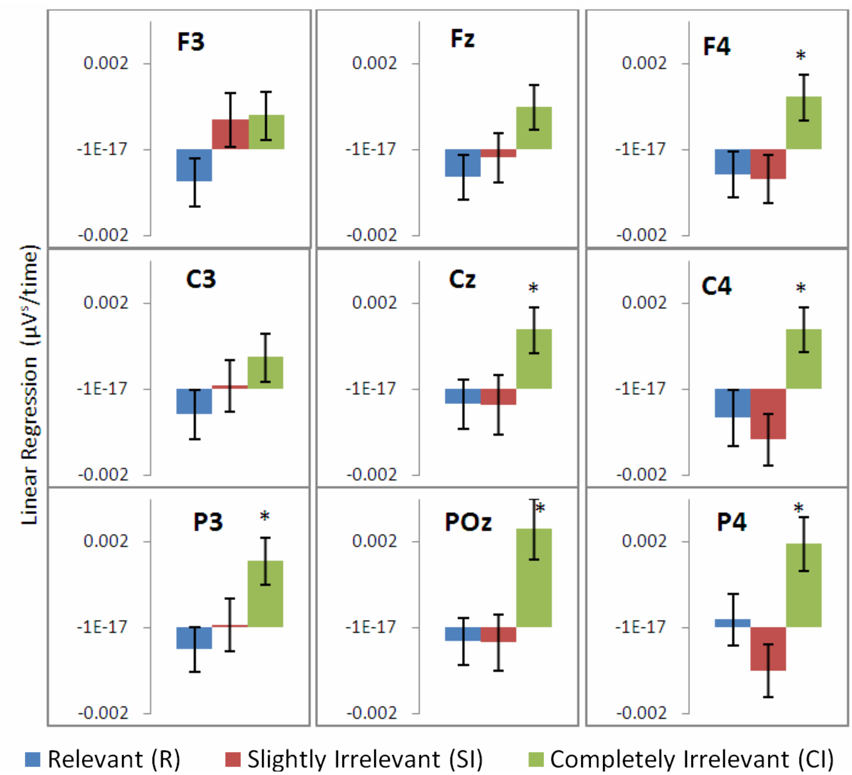


Fig. 3. Slope of regression lines of the Fast Theta band during processing of R, SI, and CI sentences, n=9. Measurement began at time of sentence presentation; ended at keyed response.

Significant effects were also found in the slopes of regression lines of slow alpha (8-10 Hz) during the first second of sentence processing (Fig. 4). ANOVA tests revealed significant differences at channels F4 [$F\ (2,370) = 6.194$, $p < 0.01$] and C4 [$F\ (2,370) = 5.218$, $p < 0.01$]. Post hoc comparisons found that in channels F4 and C4, the mean linear regression of CI sentences is significantly different from both R and

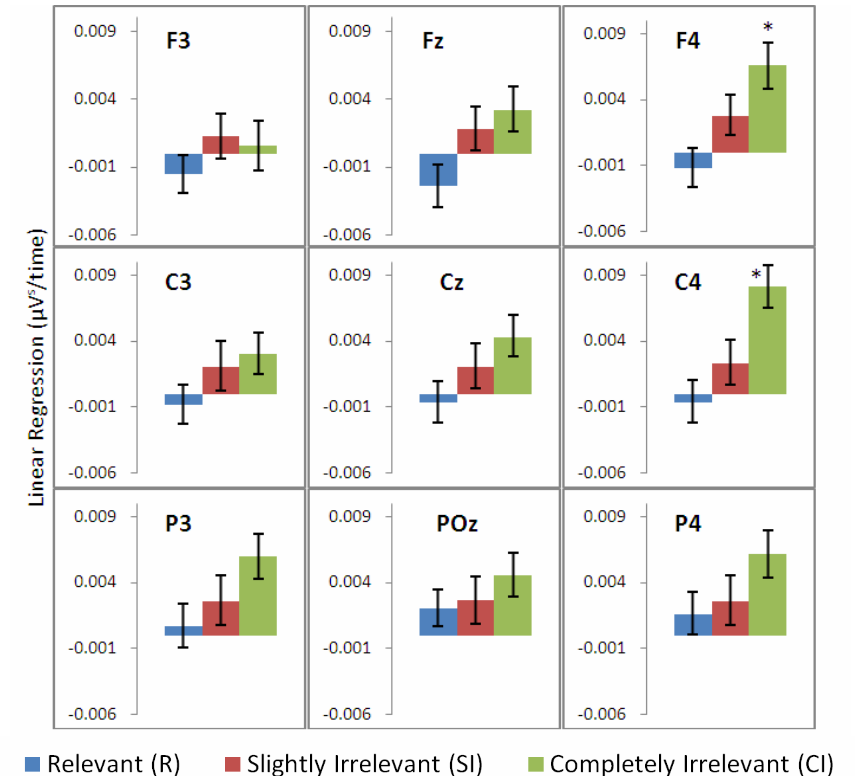


Fig. 4. Slope of regression lines of the Slow Alpha band during processing of R, SI, and CI sentences, n=9. Measurement began at sentence presentation; ended one second later, regardless of when response occurred.

SI sentence types, although R and SI are not different from each other. These findings match the right-lateralized pattern found in fast theta during the entire duration of sentence processing. However, contrary to the fast theta findings, the increase in slow alpha for CI sentences is stronger in the frontal and central regions than in the posterior region.

4 Discussion

Previous research suggests that phasic (event-related) changes in theta and alpha EEG frequencies reflect different types of cognitive processing [8-11]. Theta band synchronization (increase in power) is positively correlated with the encoding of new information, i.e. working memory or episodic memory in particular [12], while alpha band desynchronization (decrease in power) has been suggested as indicative of greater attentional demands [10].

The current findings suggest that the processing of relevant compared to completely irrelevant items causes differential changes in the EEG theta and alpha bands, particularly fast theta (5-7 Hz) and slow alpha (8-10 Hz). The significantly greater increase in power of the fast theta frequency band during the processing of CI sentences over that of R or SI sentences most likely reflects the encoding of novel information [8-11]. CI sentences contain items of random, non-relevant facts that could not be fit into the pre-existing mental model that represents the context of the analysis question. For example, in a scenario regarding suspicious Al-Qaeda activity, a CI sentence read, “Most children prefer cookies to vegetables, unfortunately for their health-conscious parents.” Therefore we propose that the processing of these new, ‘out-of-context’ snippets of information cause the fast theta band to increase in power. The opposite pattern is seen in the processing of relevant items. At all sites other than P4, R sentences produce a decrease in fast theta power. This is most likely due to the matching of this information with the participants’ mental model of the analysis context. SI sentences show a pattern most similar to that of relevant items. The ‘key words’ contained in SI sentences likely cause the participant to immediately try to fit the new piece of information within the context of the analysis question. Therefore the same novel response as seen in the CI sentences is not observed. It should be noted that the similarity of the SI and R EEG patterns indicate a risk of False Alarms with the current assessments, and this issue will be addressed as we continue to develop the system.

The localized patterns of fast theta response are also of interest. The slope of regression lines for CI is most positive in the right-hemisphere sites, and at the posterior sites. The greatest increase in fast theta power for CI is found at POz. The right-hemisphere lateralization may reflect the content of the CI sentences that were often novel, surprising and even humorous in the context of the serious nature of the material deemed relevant to the hypotheses. The decrease in fast theta during the processing of R sentences is most negative in the left hemisphere, and at frontal sites. The F3 electrode site shows the greatest decrease in fast theta for items of relevance. It is possible that the left lateralized theta decrease is related to the perceived match to the mental model, thus reducing the requirement for further analysis of the context or the semantics of the sentence.

Responses in the slow alpha band (8-10 Hz) are similar to those observed for fast theta: slow alpha power increased significantly more for CI sentences than for R or SI sentences. In fact, slow alpha power for R sentences decreased at most electrode sites. Our findings (alpha desynchronization for relevant sentences) are consistent with previous research indicating that demanding and/or relevant tasks are associated with a greater level of alpha desynchronization than less demanding and/or irrelevant tasks [13]. Additionally, significant differences between CI and SI sentences and R sentences are strongest on the right hemisphere at electrode sites F4 and C4. Increased alpha power on the right hemisphere for CI and SI sentences may represent increased attention to a surprising or novel event. Boiten (1991) [11] reported an increase in alpha power in the right hemisphere that occurred when an input was surprising, complex or novel, which in our experiment could represent the presentation of CI (and to an extent SI) sentences.

It is important to note that while our study documents significant differences between CI, SI and R sentences for both fast theta and slow alpha levels, there is still a large amount of inter-individual variance. Fast theta and slow alpha bands were selected to represent 5-7 Hz and 8-10 Hz respectively. However, these bandwidths are arbitrary and optimal theta and alpha ranges may differ between individuals. Age, brain volume, neurological disorders, education, memory performance etc., all influence peak alpha and theta frequencies [12-14]. Our fixed frequency bands may in fact be an intermingling of “true” alpha and theta for each individual, and thus may not accurately isolate the physiological effects of processing R, SI and CI sentences. In other words, our “Fast Theta” may actually represent some slow alpha activity and vice versa. For future studies it may be necessary to define the alpha and theta bands individually for each subject, in a similar manner as that described by Klimesch [13], in order to accurately isolate the neurophysiological components of processing R, SI and CI sentences in single trials.

Though preliminary, these findings highlight the role of task demands and task relevancy in triggering changes in the EEG theta and alpha bands. The data supports the concept of an automated relevancy indicator based on neurophysiological responses. By monitoring the evidence gathering process through neurophysiological sensors and implementation of real-time strategies, more accurate and efficient extraction of evidence may be achieved.

Acknowledgements

This work was supported by DARPA Contract No. W31P4Q-06-C-0041.

Approved for Public Release, Distribution Unlimited.

References

1. Welty, C., Murdock, J.W., Pinheiro da Silva, P., McGuinness, D., Ferrucci, D., Fikes, R.: Tracking information extraction from intelligence documents. In: 2005 International Conference on Intelligence Analysis (IA 2005), McLean, VA, USA (2005)
2. Cowell, D.A.: Social and Collaborative Environments Lead at Pacific Northwest National Laboratory operated by Batelle for US DOE (2006)
3. Pirolli, P., Card, S.: The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis. In: 2005 International Conference on Intelligence Analysis, McLean, Virginia (2005)
4. Jasper, H.H.: The ten-twenty electrode system of the International Federation. Electroencephalography and Clinical Neurophysiology 10, 371–375 (1958)
5. Hughes, F.J., Schum, D.A.: Evidence Marshaling and Argument Construction – Case Study #1: The Case of General Alpha: Facing Insurgency? Student Assignment, Joint Military Intelligence College, Defense Intelligence Agency (2006a)
6. Hughes, F.J., Schum, D.A.: Evidence Marshaling and Argument Construction – Case Study #2: The Case of Wigmore vs. Al-Quaeda. Student Assignment, Joint Military Intelligence College, Defense Intelligence Agency (2006b)
7. Berka, C., et al.: Real-time Analysis of EEG Indices of Alertness, Cognition, and Memory Acquired with a Wireless EEG Headset. International Journal of Human-Computer Interaction 17(2), 151–170 (2004)
8. Bastiaansen, M., Berkum, J., Hagoort, P.: Event-Related Theta Power Increases in the Human EEG During Online Sentence Processing. Neuroscience Letters 323, 13–16 (2002)

9. Klimesch, W., Doppelmayr, M., Russegger, H., Pachinger, T.: Theta band power in the human scalp EEG and the encoding of new information. Neuroreport 7, 1235–1240 (1996)
10. Fink, A., Grabner, R.H., Neuper, C., Neubauer, A.C.: EEG alpha band dissociation with increasing task demands. Cognitive Brain Research 24, 252–259 (2005)
11. Boiten, F., Sergeant, J., Geuze, R.: Event-related desynchronization: the effects of energetic and computational demands. Electroencephalogr. Clin. Neurophysiol. 82, 302–309 (1992)
12. Klimesch, W., Schimke, H., Pfurtscheller, G.: Alpha frequency, cognitive load and memory performance. Brain Topography 5, 1–11 (1993)
13. Klimesch, W.: EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res. Rev. 29(2-3), 169–195 (1999)
14. Nunez, L.R., Bickford, R.G.: The relationship of head size to alpha frequency with implications to a brain wave model, Electroencephalography. Clin. Neurophysiol. 44, 344–352 (1978)

Affective Computer-Generated Stimulus Exposure: Psychophysiological Support for Increased Elicitation of Negative Emotions in High and Low Fear Subjects

Christopher G. Courtney[1,3], Michael E. Dawson[1], Anne M. Schell[2], and Thomas D. Parsons[3]

[1] Department of Psychology, University of Southern California
[2] Department of Psychology, Occidental College
[3] Institute for Creative Technologies, University of Southern California
Courtney@ict.usc.edu, mdawson@usc.edu, schell@oxy.edu, TParsons@ict.usc.edu

Abstract. The present study examined physiological measures of affect when viewing images from the International Affective Picture System (IAPS), computer-generated still images, and computer-generated videos of feared and non-feared stimuli. Twenty low fear (LF) and twelve high fear (HF) individuals viewed static and moving images of spiders and snakes. In both LF and HF subjects, computer-generated video images elicited more intense affective responses than the IAPS images and the computer-generated stills. Computer-generated still images were as effective in eliciting fear responses as the IAPS. These results suggest that computer-generated images can be as or more effective as the IAPS in eliciting fear. Regardless of modality, HF subjects showed stronger physiological responses to their specifically feared stimulus (snake or spider) than to a non-feared stimulus.

Keywords: Psychophysiology, Fear, EMG, skin conductance, VR, startle.

1 Introduction

Fear is essential to the survival of organisms. From an evolutionary perspective, fear was developed in order to facilitate an organism's response to threat. It motivates escape from, and avoidance of, dangerous stimuli in the natural world. Accordingly, much empirical attention has been focused on deconstructing the mechanisms through which humans process and react to fear-producing stimuli.

1.1 Elicitation of Fear in the Laboratory

The study of fear responses in the laboratory has often used emotion laden pictures, such as the International Affective Picture System (IAPS, [1]). The IAPS consists of pictures meant to evoke negative, positive, or neutral affect. This collection of pictures has become the standard in psychological studies of emotion. The IAPS offers many advantages, including extensive normative data and evidence of stability across laboratories in different countries [2].

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 459–468, 2009.

Although the IAPS has been instrumental in the study of affect, preliminary evidence suggests that virtual reality (VR) stimuli may be more effective at eliciting emotional responses. VR technology allows subjects to be immersed in a three-dimensional (3D) virtual environment (VE) in which they are free to look around and explore. This may create a greater sense of presence, or feeling of “being there” [3].

Indeed, VEs may elicit arousal responses comparable to those evoked by in vivo exposure to real world stimuli. For example, Emmelkamp et al. [4] compared the responses of acrophobics randomly assigned to exposure treatment with VR and real stimuli. Subjects in the two conditions evidenced similar subjective ratings of anxiety. This suggests that VR technology can be used to create realistic environments that are as effective in eliciting fear responses as real-world stimuli. Indeed, a recent meta-analysis of VR exposure therapy outcomes concluded that it is effective in reducing phobia and anxiety symptoms [5].

However, the development of VR systems is still quite costly. Before investing the resources necessary for the development of VR stimuli, it is important to assess whether computer-generated (CG) images and videos are as effective in eliciting fear responses as images of real stimuli. Jang et al. [6] measured psychophysiological responses including skin resistance and heart rate variability to assess arousal levels in normal subjects exposed to fear of driving and fear of flying VEs. Subjects showed lowered levels of skin resistance compared to baseline, indicating higher levels of arousal. This suggests that VEs can be physiologically arousing, but direct comparisons of VR stimuli to standard pictures in a controlled within-subjects paradigm is still lacking.

In the current study, we assessed whether VR still images and videos of virtual spiders and snakes are threatening enough to elicit fear responses that are similar to or greater than those elicited by photographs of real spiders and snakes. IAPS slides were used as comparisons because they are well-validated and widely-used in the study of human affect. A within-subjects design was used to control for individual variability in responding. We hypothesize that CG stimuli will be as or more effective than IAPS images in eliciting fear responses.

VR systems are typically equipped with head tracking capabilities to allow the subject to explore his/her environment. However, IAPS slides do not lend themselves well to this type of presentation. Thus, immersive VEs were not used in this study. Instead, CG stimuli that could be used in a VE were projected onto a screen in front of the subjects to achieve greater control of what the subject was viewing, and to prevent the VR stimuli from having an advantage in creating an arousing situation due to the novelty of the head tracking capabilities. In addition to investigating the differences between CG and real stimuli, the current study sought to compare fear responses elicited by static versus moving images. The breadth of literature that has examined differences in psychophysiological responses to moving and static emotional stimuli is quite limited; however, Detenber, Simons, and Bennett [7] showed that participants exhibited stronger skin conductance and heart rate responses to moving, rather than static, images.

Moving images may be more physiologically arousing because humans may have an innate tendency to attend to moving over stationary objects. Franconeri and Simons [8] postulated that humans may have an innate tendency to attend to moving stimuli because they signal an event that could require urgent action. Thus, we hypothesize that CG videos will be more effective in eliciting fear responses than either the CG stills or IAPS images.

The current study also aimed to understand how these stimuli affect subjects who are high in fear of a specific stimulus, and to assess which stimulus modalities work best for differentiating a specifically feared stimulus and a generally fear-relevant stimulus. Responses of a high-fear (HF) group consisting of subjects who scored high on scales of fear of spiders or snakes, but not both, was compared to responses of a low-fear (LF) control group. Previous research has shown that specific phobics tend to be most responsive to their specifically feared stimuli [9]. The inclusion of the HF group allows assessment of which stimulus modality is best for targeting emotional responses to a specifically feared stimulus. This information may have immediate clinical relevance, as there is growing interest in VR exposure as a treatment for phobias.

1.2 Physiological Components of Fear

In measuring fear responses, it is important to consider physiological indices in order to gain a more complete understanding of the response. Self-report data are highly susceptible to influences outside the subject's own targeted attitudes [10]. Emotional responses are commonly thought of as varying across two dimensions: valence and arousal. In this study, we use skin conductance responses as an index of arousal, whereas startle eyeblink responses were used as an index of valence.

Skin conductance responses provide an atypical and useful index of autonomic functioning in that they are mediated exclusively by the sympathetic nervous system. While skin conductance provides a reliable measure of arousal, or motivational intensity [11], it is not an optimal method for differentiating between appetitive and avoidance motivation. Therefore, we also employed electromyographic (EMG) recordings of the startle eyeblink reflex, a widely used psychophysiological index of valence. Vrana et al. [12] found that startle responses are facilitated when presented in conjunction with a negative stimulus, and inhibited when presented with a positive stimulus relative to presentations with neutral stimuli, an effect found to be highly reliable [13].

1.3 Hypotheses

Our primary objective was to examine the effectiveness of IAPS slides, CG stills, and CG videos in eliciting fear responses in HF and LF subjects. We expected that, across groups, CG videos would elicit the highest levels of arousal, as indexed by skin conductance responses. Moreover, CG videos were hypothesized to produce the strongest potentiation of startle eyeblink. We further hypothesized that CG still images would be as effective as IAPS slides in eliciting skin conductance responses and eyeblink potentiation. A second set of goals of the current project consisted of investigation of the physiological responses to feared versus non-feared stimuli among HF subjects. We expected that HF subjects would show stronger responses to their feared stimulus (e.g., snake) than to a non-feared stimulus (e.g., spider).

2 Methods

2.1 Participants

Thirty-two participants (22 females, mean age = 20.59) were selected based on a questionnaire screening of 407 college students. Participants were selected based on their scores on the Spider Questionnaire (SPQ) and Snake Questionnaire (SNAQ; see [14]). Participants were selected for the HF group if their scores were above the 90th percentile on either the SPQ or the SNAQ and below the mean on the other questionniare. The HF group consisted of 12 participants (8 spider fearing and 4 snake fearing). The LF group was selected to match the range of the HF group's scores regarding the non-feared object.

2.2 Stimuli and Design

Participants viewed snakes and spiders using three different media types, including pictures taken from the International Affective Picture System [1], computer-generated (CG) videos, and CG still pictures. Each stimulus was projected onto a screen (33 inches high, 41 inches wide) for five seconds with a 15 to 20 second inter-trial interval.

Four IAPS pictures of both snakes and spiders were selected. Valence and arousal ratings were similar across animal types [1].

Video clips with 3D graphic virtual reality content of four snakes and four spiders were first storyboarded and designed on paper, and then models were built in Maya before being converted to OpenGL models. Spiders and snakes varied in shape, form, and size. Clips also differed in background environment.

Four CG still-framed pictures of both spiders and snakes were taken from the CG videos. Still images were selected in an attempt to match the way each animal was presented in the IAPS pictures, and were considered representative of the video from which each was derived .

The experimental test session consisted of eight blocks of six trials each. A five-minute break followed the first four blocks. During this break, subjects filled out a demographics questionnaire. Block presentation order was counterbalanced across subjects. Each block consisted of one snake and one spider from each of the three media types. The six types of stimuli (snake IAPS, snake CG still, snake CG video, spider IAPS, spider CG still, and spider CG video) were counterbalanced to appear in each ordinal position within the blocks the same average number of times, and each stimulus type was presented before and after each other stimulus type the same average number of times. Each stimulus was presented exactly once during the first four blocks and once during the second four blocks. The deleted information was provided above.

An acoustic startle-eliciting stimulus was presented during three of the six trials of each block. The startling stimulus was not presented during the same type of stimulus on consecutive blocks and no more than three consecutive trials included a startling stimulus. A total of 24 startling stimuli were presented in the experiment. The startle eliciting stimulus was a 110 dB white noise burst 50 ms in duration with a near instantaneous rise/fall time presented binaurally through Telephonics TDH-50P headphones. Decibel levels were measured with a Realistic sound level meter using a Quest Electronics earphone coupler.

2.3 Dependent Variables

Electromyographic (EMG) and skin conductance responses (SCRs) were recorded simultaneously throughout the experiment using Contact Precision Instruments equipment and a computer running SAM1 software.

Startle eyeblink response. EMG startle eyeblink responses were recorded using two miniature silver-silver chloride electrodes (4 mm in diameter) placed over the orbicularis oculi muscle of the left eye. One electrode was placed directly below the pupil in forward gaze while the other was placed about 1 cm lateral to the first. A large silver-silver chloride electrode (8 mm in diameter) was placed behind the left ear to serve as a ground.

The raw EMG signal was recorded at a rate of 1000 Hz throughout the experimental session using a 10 Hz high pass and 200 Hz low pass filter. Startle responses were rectified and integrated for analysis using a 20 ms time constant. In order to be scored, the onset of the blink response had to occur within a window of 20 to 100 ms following the startle probe. The blink response had to reach peak activity within a window of 20 to 150 ms following the startle probe. Amplitudes were recorded as the difference between the peak activity value and the baseline level that was present immediately preceding onset of the blink response. Subjects who failed to reach 1 μV amplitudes on more than 50% of startling trials were considered non-responders and were dropped from further EMG analyses. Two subjects from the LF group reached this criterion. If the subject was blinking during the onset of the startle stimulus, the blink was removed from scoring due to artifact. These blinks were replaced with the average of that subject's blinks to the other three startled trials of that stimulus type (i.e., the same animal and modality). Outliers were defined as being 3 standard deviations above the mean for each subject as well as being 2 standard deviations above the next largest response from that subject. Only one response from one subject was determined to be an outlier in the current study and was replaced using the same methods used to replace blinks removed due to artifact.

Due to the high levels of variability between subjects in EMG responses, and because of a relatively small sample size in the HF group, all blink amplitude values were standardized using a within subject z-transformation. This helped to ensure that all subjects contributed to group means equally.

Skin conductance response. SCR was measured with the use of 8 mm silver-silver chloride electrodes placed on the volar surface of the distal phalanges of the index and middle fingers of the non-dominant hand. Electrodes were filled with a 0.05 molar isotonic NaCl paste to provide a continuous connection between the electrodes and the skin.

Skin conductance responses were scored as the largest amplitude response beginning in a window of 1 to 3 s following stimulus onset. A response was defined as having a peak amplitude greater than 0.01 μS.

3 Results

For each dependent variable a mixed analysis of variance (ANOVA) was run with a 3 (media) by 2 (animal type) within subjects design and a group (HF vs. LF) between subjects factor. Another set of ANOVAs was run for both the HF group and the LF control group separately. In the ANOVA for the HF group, spiders and snakes were categorized as being either a "feared" stimulus, meaning those stimuli that the subject was specifically afraid of (i.e. snakes for snake-fearing subjects), or a "non-feared" stimulus, that is a stimulus that the subject was not specifically afraid of, but which may be biologically prepared to be fear inducing (i.e. spiders for snake-fearing subjects). A 3 (media) by 2 (stimulus fear level) repeated measures ANOVA was run to examine the effects of these variables on the HF group. A separate 3 (media) by 2 (animal type) repeated measures ANOVA was run for the LF group in order to check for possible differences in responding to the different animal types (snakes versus spiders) and media. All significant media main effects were followed up with paired samples t-tests in order to identify the precise nature of these effects. A Bonferroni correction for multiple comparisons procedure was used to prevent inflation of type 1 error rates [15]. All significant t-test results reported here are Bonferroni corrected.

3.1 EMG Results

In the ANOVA involving all subjects, an overall media main effect was found, F (2, 30) = 27.85, $p < 0.001$. This effect was the result of larger eyeblink responses during CG video stimuli. Responses to CG video presentations were significantly larger than response to IAPS , t (29) = 5.578, $p < 0.01$, and CG stills, t (29) = 7.946, $p < 0.01$. Responses to IAPS and CG stills did not differ significantly ($p = 0.23$).

The analysis of the LF group alone revealed the same main effect of media, F (2, 18) = 23.648, $p < 0.001$. Responses during CG video presentations were larger than those recorded during IAPS presentations, t (17) = 4.838, $p < 0.01$, or CG still presentations, t (17) = 6.892, $p < 0.01$. Startle eyeblink respnses elicited during IAPS and CG still presentations did not differ significantly (p = 0.29). An unexpected animal type main effect was also found, F (1, 18) = 5.492, $p < 0.05$. Larger responses occurred when viewing snakes than when viewing spiders (snake viewing mean standard score = 0.0987; spider viewing mean = -0.1124).

The analysis of the HF group alone revealed that HF subjects also showed a main effect of media, F (2, 12) = 8.385, $p < 0.01$ (see Figure 1). Once again, CG videos elicited larger responses than IAPS, t (11) = 2.863, $p < 0.05$, or CG stills, t (11) = 4.137, $p < 0.05$. Responses during the IAPS and CG still stimuli did not differ significantly (p = 0.51). A main effect of stimulus fear relevance was also found in HF subjects, F (1, 12) = 6.019, $p < 0.05$. Subjects had larger eyeblink responses during presentations of their specific feared stimuli (M = 0.1553) than during their non-feared stimuli (M= -0.1835), and this effect was consistent across the three media.

3.2 SCR Results

In the ANOVA of all subjects, there was a significant overall media main effect, F (2, 30) = 27.851, $p < 0.001$. SCRs elicited during CG video presentations were significantly

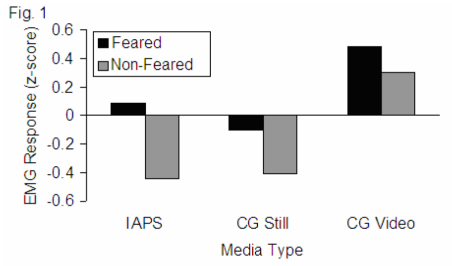


Fig. 1. High fear group's mean startle eye-blink responses to feared and non-feared stimuli across media type

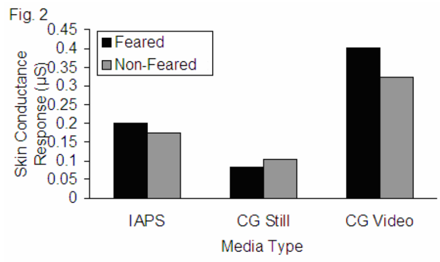


Fig. 2. High fear group's mean skin conductance responses to feared and non-feared stimuli across media type

greater than those elicited during IAPS presentations, t (31) = 3.511, $p < 0.01$, and those elicited during CG still presentations, t (31) = 3.165, $p < 0.01$. SCRs elicited during IAPS and CG still presentations did not differ significantly ($p = 0.328$).

The LF group showed a main effect of media, F (2, 20) = 5.095, $p < 0.05$. However, the pattern of responding was slightly different than that obtained when all subjects were combined. SCRs elicited during CG video presentations were again larger than those recorded during IAPS presentations, though not significantly so after Bonferroni correction, t (19) = 2.374, $p < 0.1$. SCRs elicited during CG still presentations were also greater than those elicited during IAPS presentations, t (19) = 2.641, $p < 0.05$. Videos also had a tendency to elicit greater SCRs than CG stills, though not significantly so ($p = 0.074$). Similar to the EMG results, an unexpected main effect for animal type was also found in SCRs, though this was only a trend, F (1, 20) = 3.406, $p = 0.081$. Subjects again tended to have larger responses when viewing snake stimuli ($M = 0.1378$) than spider stimuli ($M = 0.0885$).

In the HF group, only a trend toward a media main effect was found, F (2, 12) = 3.27, $p = 0.081$ (see Figure 2). Subjects tended to have larger SCRs when viewing videos ($M = 0.3623$) than IAPS ($M = 0.1876$) or CG stills ($M = 0.0936$), although these differences in responding were not significant after Bonferroni correction. A paired t-test also revealed a trend of increased responses in feared versus non-feared during the CG videos, t (11) = 1.809, $p = 0.098$.

4 Discussion

4.1 Effects of Media Type on Fear Responses

Previous studies of fear responses have typically relied on static pictorial stimuli such as the IAPS. In the current study, we sought to examine the effectiveness of moving images in eliciting physiological measures of affect. Consistent with our hypotheses, CG video moving stimuli were more arousing than the CG and IAPS still images, as measured by the skin conductance responses. When compared to still images, video stimuli also exhibited more negative valence, as indicated by greater startle eyeblink responses. These findings suggest that, compared to still images of real objects, VR-style stimuli can be more effective in instigating arousal in both low fear and high fear subjects.

Across LF and HF groups, subjects displayed increased SCRs to CG video stimuli compared to CG still and IAPS pictures. The motion component of video stimuli may elicit greater arousal by creating a stronger sense of presence, or "being there," in participants. Although presence is typically discussed in relation to virtual environments, it has also been used in the context of other media forms, such as television and film [16]. In the current study, we found a strong effect of media on physiological arousal.

Startle eyeblink responses, which are thought to be a sensitive psychophysiological measure of valence [12], were affected by the media manipulation in the same way as the SCRs, which are primarily sensitive to arousal [17]. One explanation for these seemingly contradictory findings is that startle eyeblinks are potentiated when viewing negatively valenced images, but these effects are augmented by arousal [11]. It follows that if all stimuli are negatively valenced, as in the current study, it will be arousal that contributes to differential responding to different stimuli.

Psychophysiological responses to IAPS and CG still images were also consistent with our hypotheses. Startle eyeblink responses and SCRs did not differ when subjects viewed IAPS and CG still images, suggesting that motion was the key factor in eliciting increased responding. These findings also indicate that VR style stimuli can be as or more effective than still images of real stimuli, such as the pictures of the IAPS, in instigating fear responses.

4.2 Effects of Animal Type on Fear Responses in LF Subjects

While the media main effect followed the expected pattern in LF subjects, an unexpected main effect of animal type was also present. Snakes elicited larger SCRs than spiders. Snakes also elicited larger eyeblink amplitudes overall, though this was mainly due to highly differential responses elicited during the CG videos, which in turn led to a media by animal type interaction. These results were unexpected because part of the selection criteria for LF subjects included having very similar scores on the SNAQ and SPQ assessments of snake and spider fear, respectively. Differences in unexplored features of the videos may account for this discrepancy. For example, two of the snake videos involved significant camera movement in addition to movement of the snake, whereas the background was relatively stationary in the spider videos. The present findings call for further investigation of the effects of different feature aspects of video presentations of affective stimuli.

4.3 Effects of Stimulus Fear Level in HF Subjects

Results for the HF group confirm hypotheses that startle eyeblink responses were more pronounced for feared than for non-feared stimuli. HF subjects responded with increased startle eyeblink responses when viewing their feared stimuli, as compared to non-feared stimuli. Surprisingly, skin conductance was not as sensitive to different levels of stimulus fear in HF subjects. HF subjects only displayed a trend toward larger SCRs in responses to the feared stimuli in the CG videos. This pattern of skin conductance responding may be a result of the possible confound of snakes eliciting higher arousal responses in general in this study, and the presence of more spider fearful than snake fearful subjects in the HF group meant that snakes were usually the non-feared stimulus. The LF group also had increased SCRs to snakes compared to spiders. While these findings will need to be replicated, the trend toward differential responding between stimulus fear levels in CG videos does suggest that the CG videos are more effective in fear instigating for specifically feared objects.

4.4 Conclusion

To our knowledge, this is the first study to examine fear responses to CG videos, stills, and IAPS images. One of the main goals of the present research was to provide validity for the effectiveness of VR stimuli in eliciting fear responses in HF subjects. Results suggest that VR stimuli can be as effective, or more effective, than pictures of "real" stimuli, even when viewed on a two-dimensional screen. These findings suggest that VR may be useful in the clinical assessment and treatment of phobias. Future research examining fear responses to CG images and videos in an immersive VE may further validate the effectiveness of VR stimuli in the study of human defensive systems.

References

1. Lang, P.J., Bradley, M.M., Cuthbert, B.N.: International Affective Picture System (IAPS): Instruction Manual and Affective Ratings. Technical report A-6, The Center for Research in Psychophysiology, University of Florida (2005)
2. Hamm, A.O., Schupp, H.T., Weike, A.I.: Motivational organization of emotions: Autonomic changes, cortical responses, and reflex modulation. In: Davidson, R.J., Scherer, K.R., Goldsmith, H.H. (eds.) Handbook of Affective Science, pp. 596–618. Oxford University Press, Oxford (2003)
3. Witmer, B.G., Singer, M.J.: Measuring Presence in Virtual Environments: A Presence Questionnaire. Presence 7, 225–240 (1998)
4. Emmelkamp, P.M.G., Krijn, M., Hulbosch, A.M., de Vries, S., Schuemie, M.J., van der Mast, C.A.P.G.: Virtual Reality Treatment Versus Exposure In Vivo: A Comparative Evaluation in Acrophobia. Behaviour Research and Therapy 40, 509–516 (2002)
5. Parsons, T.D., Rizzo, A.A.: Affective Outcomes of Virtual Reality Exposure Therapy for Anxiety and Specific Phobias: A Meta-analysis. J. of Behavior Therapy and Experimental Psychiatry 39, 250–261 (2007)
6. Jang, D.P., Kim, I.Y., Nam, S.W., Wiederhold, B.K., Wiederhold, M.D., Kim, S.I.: Analysis of Physiological Response to Two Virtual Environments: Driving and Flying Simulation. CyberPsychology and Behavior 5, 11–18 (2002)
7. Detenber, B.H., Simons, R.F., Bennett, G.G.: Roll 'em!: The Effects of Picture Motion on Emotional Responses. J. of Broadcasting and Electronic Media 42, 113–127 (1998)
8. Franconeri, S.L., Simons, D.J.: Moving and Looming Stimuli Capture Attention. Perception and Psychophysics 7, 999–1010 (2003)

9. Öhman, A., Soares, J.J.F.: "Unconscious Anxiety": Phobic Responses to Masked Stimuli. J. of Abnormal Psychology 103, 231–240 (1994)
10. Schwarz, N.: Self-reports: How the Questions Shape the Answers. American Psychologist 54, 93–105 (1999)
11. Lang, P.J.: The Emotion Probe: Studies of Motivation and Attention. American Psychologist 50, 372–385 (1995)
12. Vrana, S.R., Spence, E.L., Lang, P.J.: The startle Probe Response: A New Measure of Emotion? Journal of Abnormal Psychology 97, 487–491 (1988)
13. Bradley, M.M., Cuthbert, B.N., Lang, P.J.: Affect and the Startle Reflex. In: Dawson, M.E., Schell, A.M., Bohmelt, A.H. (eds.) Startle Modification: Implications for Neuroscience, Cognitive Science, and Clinical Science, pp. 157–183. Cambridge University Press, Cambridge (1999)
14. Klorman, R., Weerts, T.C., Hastings, J.E., Melamed, B.G., Lang, P.J.: Psychometric Descriptions of Some Specific-fear Questionnaires. Behavior Therapy 5, 401–409 (1974)
15. Holm, S.: A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics 6, 65–70 (1979)
16. Dillon, C., Keogh, E., Freeman, J.: "It's Been Emotional": Affect, Physiology, and Presence. In: Proceedings of the Fifth Annual International Workshop on Presence, Porto, Portugal (2002)
17. Dawson, M.E., Schell, A.M., Filion, D.L.: The electrodermal system. In: Cacioppo, J.T., Tassinary, L.G., Berston, G. (eds.) Handbook of Psychophysiology, pp. 159–181. Cambridge University Press, Cambridge (2007)

Applying Real Time Physiological Measures of Cognitive Load to Improve Training

Joseph T. Coyne[1], Carryl Baldwin[2], Anna Cole[3],
Ciara Sibley[2,3], and Daniel M. Roberts[2]

[1] Naval Research Laboratory, Washington, DC 20375
[2] George Mason University, Department of Psychology, Fairfax, VA 22030
[3] Strategic Analysis Incorporated, Arlington, VA 22203
cbaldwi4@gmu.edu

Abstract. This paper discusses how the fields of augmented cognition and neuroergonomics can be expanded into training. Several classification algorithms based upon EEG data and occular data are discussed in terms of their ability to classify operator state in real time. These indices have been shown to enhance operator performance within adaptive automation paradigms. Learning is different from performing a task that one is familiar with. According to cognitive load theory (CLT), learning is essentially the act of organizing information from working memory into long term memory. However, our working memory system has a bottleneck in this process, such that when training exceeds working memory capacity, learning is hindered. This paper discusses how CLT can be combined with multiple resource theory to create a model of adaptive training. This new paradigm hypothesizes that a system that can monitor working memory capacity in real time and adjust training difficulty can improve learning.

1 Introduction

Whether it is called the field of Augmented Cognition or Neuroergonomics [1] there has been a recent push to apply findings from the field of neuroscience and the "Decade of the Brain" to improve human performance. While Neuroergonomics focuses on how neuroscience can be applied to work and everyday environments, Augmented Cognition places an emphasis on the design of closed loop systems based upon real time physiological assessment. Until recently there has not been a focus on how advances within these two areas could be applied to learning or training. As the use of computers and simulation become an increasingly important component of the learning process, it seems that applying these two fields and developing a closed loop adaptive training system would be a natural extension of these two areas. Such a system could adjust the content, presentation format, and pace of training to match the specific skills and abilities of the trainee. It is proposed that such a system would reduce the amount of time required to train an individual by reducing the amount of time the trainee is under or overloaded. The present paper reviews previous research in neuroscience and learning which serve as a theoretical basis for the adaptive training system. The paper will discuss the research done with respect to real time

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 469–478, 2009.

physiological assessment, describe the cognitive load theory and explain how these two separate research areas can be merged into a theory of adaptive training.

2 Real Time Physiological Assessment

2.1 Sensors

At the heart of neuroscience are the tools available to measure activity within the brain. There are a number of different sensor technologies available which provide either direct or indirect indices of the brain's activity. These include measures taken from the brain such as electroencephalogram (EEG), functional near-infrared imaging (fNIR), magnetoencephalography (MEG), functional magnetic resonance (fMRI), as well as indirect physiological measures of brain activity such as cardiorespiratory activity, for example heart rate (HR) and heart rate variability (HVR), as well as measures of electrodermal activity such as skin conductance and galvanic skin response (GSR) and pupilometry. The advantages of each of these methods can be assessed along three criteria including spatial resolution, temporal resolution, and ease of use [2]. Of these sensor technologies EEG has been the most widely used for real time assessment due to its temporal resolution and ease of use. Although eye tracking data only provides an indirect measure of brain activity; it is widely used, unobtrusive (particularly with new off the head systems) and easy to collect. For those reasons the present paper will focus on advances made with respect to these two different sensors.

2.2 Real Time Cognitive State

For a physiological sensor to be useful it must be sensitive to an aspect of operator state that has cognitive or performance implications (i.e., stress, arousal, mental workload). Ultimately, much of the research has been looking to find a gauge of mental workload: how hard the brain is working at a given point in time. One of the key aspects of mental workload is the relationship between the task being performed and an individual's limited pool of resources available. Wickens' multiple resource theory [3] distinguishes between three orthogonal resource dimensions including perceptual modality (e.g., visual, auditory), information code (e.g., verbal, spatial) and processing stage (e.g., encoding, central processing, and responding). Multiple resource theory parallels that of Baddeley's model of working memory [4] which also includes a spatial and verbal component. In fact there is some suggestion that resource theory is essentially synonymous with working memory [5]. Mental resource capacity can be conceptualized as essentially how much information can be maintained and manipulated in working memory. Theoretical conceptions of both constructs make the assumption that capacity is in some way limited and that a given task or set of tasks can exceed that capacity. Individual differences in working memory capacity are consistently found and these differences are strongly correlated with performance on a number of different cognitive tasks [6-8]. Because working memory capacity affects both the difficulty of and the strategies used for learning complex tasks as well as the susceptibility to different forms of distraction [9-11], its assessment may provide a powerful tool for improving existing training protocols.

The agreement between the two theories is that an individual's mental capacity is in some way limited and that a given task or set of tasks can exceed that capacity. A cognitive state gauge that provides a real time indicator of an individual's available resources could provide great insight into performance, task design and training. However, moving from a physiological signal filled with noise to a real time gauge requires a significant amount of signal processing.

2.3 EEG Algorithms

A cognitive state gauge that provides a real time indicator of an individual's current level of engagement and the availability of different types of resources could provide great insight into performance, task design and training. However, improving methods of signal amplification, filtering, and the analog to digital conversions required to extract physiological signals from the background of noise is an on-going challenge to the successful implementation of real time cognitive state gauges.

Linear Classification. Early real time metrics derived from EEG analyzed the changes of spectral power in the five frequency bands (alpha, beta, theta, gamma, and delta). These changes were used to provide an indication of the operator's engagement, attention and mental workload. Pope, Bogart, and Bartolome [12] developed an index based on a ratio of EEG power, defined as (beta/(alpha + theta)) which can be computed in real-time by calculating a running average over a 20 second window. This index was said to determine the level of engagement/alertness of an individual while performing a task. The researchers were able to demonstrate the index could be used in real time to improve performance on a vigilance task [13] and a complex tracking task [14].

A second linear algorithm for processing cognitive state from EEG is the eXecutive Load Index (XLI) [15], which was designed to monitor changes in cognitive load related to processing messages in real-time. This was done by computing the ratio of ((delta+theta)/alpha) over a moving 2 second window, with the change determined by comparing the value to the previous 20 second running average.

Researchers at Advanced Brain Monitoring (ABM) developed several gauges of cognitive state based upon linear and quadratic discriminant function analysis (DFA) [16]. The gauges for mental workload and engagement are of particular interest. The index for engagement tracks the demands for sensory processing and attentional resources, whereas the index for mental workload tracks the level of cognitive function and is considered to be a correlate of executive function. The algorithms for both indexes are derived for each individual based upon his or her EEG signals on a series of baseline vigilance tasks. The measures have both been validated in a series of basic cognitive tasks. The mental workload metric has been shown to track task demand in mental arithmetic and digit span tasks as well as show a significant correlation with subjective measures of workload and task performance. The gauge for engagement has been shown to decrease as a function of time during a vigilance task whereas workload did not. The algorithms for both engagement and workload output data every second.

Non-linear classification. Research at the Air Force Research Lab has investigated the ability of an Artificial Neural Network (ANN) to classify operator mental workload in a complex laboratory task and during a UAV simulation [17, 18]. The ANN derives its classification from EEG, EOG, and heart rate data and was successfully able to classify high versus low workload with a 85-90% accuracy rate when the ANN was trained for each individual. The ANN was also successfully implemented in an adaptive automation UAV paradigm where vehicle speed was reduced during periods of high workload. The adaptive automation system was able to significantly improve performance over both a non-adaptive system and a system with random changes to the speed. Although ANN has been shown to be highly successful, it requires a large amount of data to "train" the model. It is also unclear how stable an ANN would be for a particular individual over time.

2.4 Eye Tracking

Visual scanning strategies may provide an indication of mental workload. Di Nocera, et al. [19] implemented the Nearest Neighbor Index (NNI) to investigate whether a statistical index that provides information on dispersion of points, or fixations, would have differential patterns for high workload and low workload conditions. The index is based on the Complete Spatial Randomness (CSR) method, which is the spatial analysis equivalent of uniformly and independently distributed random variables. The index is computationally straightforward and is feasible to compute in near real-time, which lends potential to be used as a metric or trigger for adaptive training. Essentially, higher values in the NNI show higher levels of entropy in scanning. Preliminary analysis from a case study showed that higher NNI values were correlated with higher workload, and that the NNI was sensitive to varying workload conditions. However it was suggested that more studies be performed to fully understand the correlation between the randomness of fixations and mental workload.

Cognitive workload has also been evaluated using measures of eye movement and pupil dilation to detect cognitive strategy shifts [20]. A psycho-physiological index of workload based on pupil dilation, the Index of Cognitive Activity (ICA), was used in a case study by Marshall et al. to detect shifts in strategy based upon large changes of ICA. ICA does not require averaging over trials or individuals, it can be applied to a signal of any given length, and it can be computed in near real-time. ICA is calculated as the frequency of a detection of an abrupt discontinuity in the pupil signal [21]. Marshall, et al.'s study demonstrated that cognitive strategy shifts can be identified from eye tracking data, and observed fluctuations of ICA can identify the time and location of those strategy shifts. Identification of cognitive strategy shifts may be beneficial not only for instructional design based on cognitive load theory [22], but also for adaptive training.

A recent review identified and evaluated the ability of seven eye tracking metrics to classify an operator's cognitive state, while taking into account the sensitivity and specificity of the classification [23]. The metrics under evaluation included the Index of Cognitive Activity (ICA), blinks, movement, and divergence between eyes, where separate right and left eye values were calculated for the ICA, blink, and movement metrics. Each of the seven metrics can be computed in near real time, making them attractive candidates to apply and incorporate into adaptive training applications.

For statistical analysis, all metrics were transformed to a common scale ranging from 0 to 1. Two classification models, linear discriminate function analysis and non-linear neural network analysis were employed, and the sensitivity and specificity were evaluated to determine classification adequacy. Two-state classifications were calculated for three separate studies (problem solving, driving simulation, and visual search) to differentiate between an engaged or relaxed state, focused or distracted driving, and a fatigued versus alert state. For all three studies, both classification models were successful in differentiating cognitive states (69% to 92%) based solely upon the aforementioned eye metrics. Discriminant analyses with systematic elimination of each metric were conducted to confirm that all metrics were needed to obtain the same accuracy of results. In addition, it was determined that all metrics were needed to obtain the level of demonstrated classification, and that no particular metric was salient across all subjects within any study.

3 Cognitive Load Theory

Cognitive load theory (CLT) is model of learning based around components of human information processing, particularly working memory and long term memory. A core principle of CLT theory is that learning places demands on a limited capacity working memory system [24, 25]. Since working memory capacity (WMC) is limited [26, 27], learning is integrally tied to both the working memory capacity (WMC) of the learner and the working memory demand of the instruction and instructional material. While the WMC of an individual is limited, his or her long term memory is almost unlimited. Thus the theory is concerned with how information from working memory is organized and grouped together (into schemata) and stored in long term memory. Once information is stored in long term memory, it enables the individual to access it later and reduces the burden placed upon the working memory system. Much of the research on CLT has focused on working memory since it serves as the bottleneck to learning.

CLT proposes three specific types of cognitive load with additive effects; the sum total of which must not exceed a learner's working memory resources if optimal learning is to be achieved. The first, termed *intrinsic cognitive load* is the difficulty imposed by the material or task to be learned. It is heavily influenced by the elemental interactivity of the material – how many interacting elements must be maintained in working memory at any given time. Complex material may have high elemental interactivity. The more inherent elemental interactivity, the higher the cognitive load. Often there is little if anything that instructional design can do to change the intrinsic cognitive load of the material or task to be learned. As expertise develops, schemas are formed and elements become grouped together; enabling the individual to deal with more elements simultaneously and allowing them to overcome the working memory bottleneck. This process reflects learning and the dynamic nature of intrinsic cognitive load within CLT. The number of elements which make up intrinsic load are based upon the individual's ability to group them together.

The second form of cognitive load is *extraneous cognitive load* and this is where instructional design has the potential to make vast improvements in training. Extraneous load typically refers to how the information is presented, e.g., graphically versus verbally. Ineffective instructional designs impose an additional level of extraneous cognitive load, which is particularly problematic when the intrinsic load is high. Much of CLT has emphasized reducing extraneous load as a method of reducing overall load and enabling learning.

Germane cognitive load, the third type, is the process of creating and organizing information into schema. Germane load is the result of the instructional design [28]. It promotes the development of accurate mental models of the task and relevant schemas as well as facilitating the transition from controlled to automatic processing that accompanies expertise. Germane load is influenced by the manner, modality and sequence in which the material is presented and the learning activities involved. Differential sensitivity has been observed between various measures for each type of load [29].

Within the framework of cognitive load theory, intrinsic cognitive load is set by the task, and extraneous cognitive load is typically manipulated through instructional design. This ensures an individual's cognitive resources are not being exceeded, and that learning is promoted. Instructional design that reduces extraneous cognitive load frees more resources for germane cognitive load thus facilitating the development of schema acquisition and a shift toward automatic processing and expertise.

4 Adaptive Training

Although traditionally cognitive load theory has focused on adjusting extraneous load simulation based training allows for an adjustment in the amount of intrinsic cognitive load presented at a given time. This ability to manipulate intrinsic load combined with the capability to measure working memory capacity in real time provides the basis for the development of an adaptive training approach.

Combining elements from Wickens' Multiple Resource Theory [3] which is designed to describe workload and ultimately help predict performance and Sweller's CLT [24] we have developed an initial throughput model of learning that can be used in a closed loop system. Within the adaptive training paradigm the intrinsic load acts as the input into the system. How that information is presented in terms of modality and processing code produces the extraneous load. As with the traditional model of CLT presenting information in a spatial code versus a verbal code can yield different amounts of extraneous load depending on the task/information. Following multiple resource theory and CLT it would be possible to reduce extraneous load by using different modalities and processing codes.

As with CLT the germane load is still the organization of information into schema. The overall load is still the combination of intrinsic, extraneous, and germane load. However, as with multiple resource theory there are now potentially multiple capacities. Figure 1 represents a simple throughput model of how adaptive training would work. Operator capacity would be assessed in real time based upon a real-time physiological metric described above. Presently their may not be a separate metric for each potential resource pool. However an overall gauge of spare capacity could still serve to trigger an adaptive training screen. Based upon whether the physiological metrics indicate spare capacity the screen could add or remove certain elements of

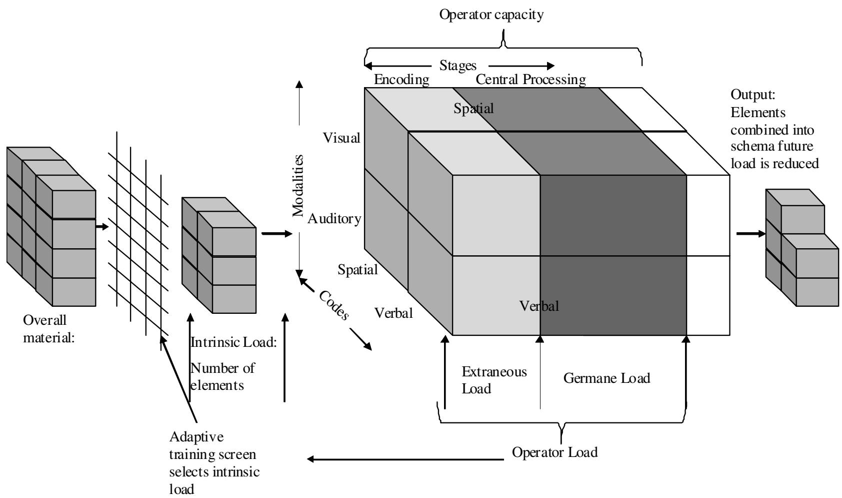


Fig. 1. Model of Adaptive Training based upon Wickens' Multiple Resource Theory and Sweller's Cognitive Load Theory

intrinsic load or add/reduce the size of a single element (e.g., driving at high speed versus low) from the overall material to be learned.

For example, in a computer simulation designed to teach target identification, the cognitive load of the task can be manipulated by adjusting the number of targets presented at a given time, target speed or salience. These manipulations do not change the intrinsic load of the task, per se. However, the load imposed on the novice learner is adjusted. Extraneous load may be reduced by making the targets more visible, by using an auditory modality to supplement identification of salient aspects of the visual targets, etc.

Germaine cognitive load, which supports skill development, can be increased by presenting various targets to be identified in random order rather than for instance all enemy tanks and then all enemy helicopters. The proposed adaptive training system would manipulate the amount of overall cognitive load in the scenario based on assessment of the trainee's current expenditure of mental resources. When physiological metrics indicate that mental workload is high, for example enemy units could slow down or decrease in number. Alternatively, when a trainee displays a low level of mental resource utilization, the scenario can be made more difficult. Changes to the cognitive load in the adaptive training system must be based not only on an individual's available resources but also their level of expertise with the system. There is a dynamic relationship between level of expertise and cognitive load required by the task.

The primary difference between a novice and an expert in any given task domain hinges on two things. First, the expert has an extensive knowledge base of well developed relevant schemas held in long term memory. Schemas allow a person to treat multiple elements as one item. For example, an expert chess player has literally thousands of schemas for movement patterns stored in long term memory. Secondly, for the expert many of the relevant tasks and skills as well as access to the stored schemas are automatic, no longer requiring resource demanding controlled processing [30]. Novices have neither extensive or well developed schemas, nor can they initiate many of the task components automatically [31]. Problem solving routines and access to schemas can become automated as when one automatically knows to solve the algebraic equations within the brackets before moving on to the relationships between the bracketed and non-bracketed items. Individual features of letters, nor even individual letters need to be processed once a reader has developed sufficient skill. As schemas develop and tasks become automated, working memory load is reduced and learning is accelerated. As learning is accelerated the amount and or rate of information presentation can *and should* be accelerated. Monitoring the transition from novice to expert is essential for efficient learning and is a key element of the proposed adaptive training strategy.

As with CLT the adaptive training paradigm will recognize the task-learner interaction, or expertise reversal effect, meaning that as a learner develops expertise, the methods of instruction that are effective should change [28, 32]. Optimal training protocols must continuously monitor in real-time both the working memory resources being utilized or mental workload of the learner and changes in skill level associated with developing expertise. Under or over utilization of working memory processes or a mismatch with the learner's current skill level will result in less efficient learning.

5 Conclusions

It is believed that the proposed adaptive training model will be able to significantly improve learning by eliminating the time in which the learner is not in an optimal state as determined by their working memory capacity. An adaptive training system will be capable of reducing the intrinsic load when working memory capacity is exceeded, or adding to the intrinsic load when there is sufficient reserve working memory capacity. Additionally the new model allows for a diagnostic approach to implementing the adaptive training screen. Advances in the real time sensors may eventually be capable of assessing capacity within the different pools (i.e., spatial versus verbal) and therefore allow for more specific changes to the material being presented. Eventually such a system will be capable of moving between information codes and processing modalities of the information being presented to capitalize on an individual's multiple resources.

References

1. Parasuraman, R.: Neuroergonomics: Research and Practice. Theoretical Issues in Ergonomics Science 4, 5–20 (2003)
2. Parasuraman, R., Wilson, G.F.: Putting the Brain to Work: Neuroergonomics Past, Present, and Future. Human Factors 50, 468–474 (2008)
3. Wickens, C.D.: Processing resources in attention. In: Parasuraman, R., Daives, D.R. (eds.) Varieties of attention, pp. 63–102. Academic Press, New York (1984)
4. Baddley, A.D.: Working Memory. Oxford University Press, Oxford (1986)
5. Parasuraman, R., Caggiano, D.: Neural and genetic assays of mental workload. In: McBride, D., Schmorrow, D.D. (eds.) Quantifying Human Information Processing Rowman and Littlefield, Lanham, MD (2005)

6. Conway, A.R.A., Engle, R.W.: Individual Differences in Working Memory Capacity: More Evidence for a General Capacity Theory. Memory 4(6), 577–590 (1996)
7. Brumback, C.R., Low, K.A., Gratton, G., Fabiani, M.: Putting Things into Perspective: Individual Differences in Working-Memory Span and the Integration of Information. Experimental Psychology 52(1), 21–30 (2005)
8. Unsworth, N., Engle, R.W.: The Nature of Individual Differences in Working Memory Capacity: Active Maintenance in Primary Memory and Controlled Search From Secondary Memory. Psychological Review 114(1), 104–132 (2007)
9. Cantor, J., Engle, R.W.: Working-memory capacity as long-term memory activation: An individual-differences approach. Journal of Experimental Psychology: Learning, Memory, and Cognition 19(5), 1101–1114 (1993)
10. Sobel, K.V., Gerrie, M.P., Poole, B.J., Kane, M.J.: Individual differences in working memory capacity and visual search: The roles of top-down and bottom-up processing. Psychonomic Bulletin & Review 14(5), 840–845 (2007)
11. Graf, S., Lin, T., Kinshuk: The relationship between learning styles and cognitive traits–Getting additional information for improving student modelling. Computers in Human Behavior 24(2), 122–137 (2008)
12. Pope, A.T., Bogart, E.H., Bartolome, D.S.: Biocybernetic system evaluates indices of operator engagement in automated task. Biological Psychology 40(1), 187–195 (1995)
13. Mikulka, P.J., Scerbo, M.W., Freeman, F.G.: Effects of a biocybernetic system on vigilance performance. Human Factors 44, 654–664 (2002)
14. Freeman, F.G., Mikulka, P.J., Scerbo, M.W., Prinzel, L.J., Clouatre, K.: Evaluation of a psychophysiologically controlled adaptive automation system, using performance on a tracking task. Applied Psychophysiology and Biofeedback 25(2), 103–115 (2000)
15. DuRousseau, D.R., Mannucci, M.A.: eXecutive Load Index (XLI): Spaital-frequency EEG tracks moment-to moment changes in high-order attentional resourcesFoundations of Augmented Cognition, pp. 245–251. Lawrence Erlbaum Associates, Mahwah (2005)
16. Berka, C., Levendowski, D., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P., Craven, P.L.: EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks. Aviation, Space, and Environmental Medicine 78, B231–B244 (2007)
17. Wilson, G.F., Russell, C.A.: Real-Time Assessment of Mental Workload Using Physiological Measures and Artifiicial Neural Networks. Human Factors 45, 635–643 (2003)
18. Wilson, G.F., Russell, C.A.: Operator Functional State Classification Using Multiple Psychophysiological Features in an Air Traffic Control Task. Human Factors 45(3), 381–389 (2003)
19. Di Nocera, F., Terenzi, M., Camilli, M.: Another look at scanpath: Distance to nearest neighbour as a measure of mental workload. In: de Waard, D., Brookhuis, K.A., Toffetti, A. (eds.) Developments in human factors in transportation, design, and evaluation, pp. 295–303. Shaker, Maastricht (2006)
20. Marshall, S.P., Pleydell-Pearce, C.W., Dickson, B.T.: Integrating psychophysiological measures of cognitive workload and eye movements to detect strategy shifts. In: Proceedings of the 36th Annual Hawaii International Conference on System Sciences, vol. 6 (2003)
21. Marshall, S.P.: The Index of Cognitive Activity: measuring cognitive workload. In: Proceedings of the 2002 IEEE 7th Conference on Human Factors and Power Plants, pp. 7-5–7-9 (2002)
22. van Gog, T., Kester, L., Nievelstein, F., Giesbers, B., Pass, F.: Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. Computers in Human Behavior (25), 325–331 (2009)
23. Marshall, S.P.: Identifying Cognitive State from Eye Metrics. Aviation, Space, and Environmental Medicine 78(5, Section II), B165–B175 (2007)

24. Sweller, J.: Cognitive load during problem solving: Effects on learning. Cognitive Science: A Multidisciplinary Journal 12(2), 257–285 (1988)
25. Sweller, J.: Discussion of Emerging Topics in Cognitive Load Research: Using Learner and Information Characteristics in the Design of Powerful Learning Environments. Applied Cognitive Psychology 20(3), 353–357 (2006)
26. Miller, G.A.: The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review 63, 81–97 (1956)
27. Kahneman, D.: Attention and effort. Prentice Hall, Englewood Cliffs (1973)
28. Paas, F., Renkl, A., Sweller, J.: Cognitive load theory and instructional design: Recent developments. Educational Psychologist 38(1), 1–4 (2003)
29. DeLeeuw, K.E., Mayer, R.E.: A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. Journal of Educational Psychology 100(1), 223–234 (2008)
30. Shiffrin, R.M., Schneider, W.: Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. Psychological Review 84(2), 127–190 (1977)
31. Marcus, N., Cooper, M., Sweller, J.: Understanding instructions. Journal of Educational Psychology 88(1), 49–63 (1996)
32. Ayres, P.: Impact of Reducing Intrinsic Cognitive Load on Learning in a Mathematical Domain. Applied Cognitive Psychology 20(3), 287–298 (2006)

Considerations for Designing Response Quantification Procedures in Non-traditional Psychophysiological Applications

A.V. Iyer[1,3], L.D. Cosand[2,3], C.G. Courtney[2,3], A.A. Rizzo[3], and T.D. Parsons[3]

[1] Department of Biomedical Engineering, University of Southern California
[2] Department of Psychology, University of Southern California
[3] Institute for Creative Technologies, University of Southern California
{aiyer,cosand,Courtney,arizzo,TParsons}@ict.usc.edu

Abstract. Psychophysiological assessment in the context of virtual environments is a promising means for benchmarking the efficacy and ecological validity of virtual reality scenarios. When applied to human-computer interaction, psychophysiological and affective computing approaches may increase facility for development of the next generation of human-computer systems. Such systems have the potential to use psychophysiological signals for user-feedback and adaptive responding. As the composition of investigating teams becomes diverse in keeping with interdisciplinary trends, there is a need to review de-facto standards of psychophysiological response quantification and arrive at consensus protocols adequately addressing the concerns of basic researchers and application developers. The current paper offers a demonstration of the ways in which such consensus scoring protocols may be derived. Electromyographic eye-blink scoring from an immersion investigation is used as an illustrative case study.

1 Background

1.1 Psychophysiology in Human Computer Interaction Research

Psychophysiology is increasingly incorporating virtual reality environments into human computer interface research [1]. The use of psychophysiological measures in studies of persons immersed in high-fidelity virtual environment scenarios offers the potential to develop current physiological computing approaches [2] into affective computing [3] scenarios. Such scenarios offer the potential for simulated environments to proffer cogent and calculated response approaches to real-time changes in user emotion, neurocognition, and motivation. The value in using virtual reality technology to produce simulations targeting emotional, neuropsychological, and motivational applications has been acknowledged by an encouraging body of research. Some of the work in this area has addressed affective processes: anxiety disorders, pain distraction and posttraumatic stress disorder [4]. Other work has assessed neurocognitive processes such as attention and executive functioning [5], [6]; memory [7], [8], [9]; and visuospatial abilities [10], [11], [12].

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 479–487, 2009.

Thus far, the recording of psychophysiological variables while participants operate within virtual environments has produced useful results in studies examining immersion and presence [13], [14], [15]. As such, the VR assets that allow for precise stimulus delivery within ecologically enhanced scenarios appears well matched for this research. One area of increasing interest in psychophysiological assessment of persons immersed in virtual environments is startle eyeblink modification (SEM) [16], [17]. The modulatory influence of selective attention as well as generalized arousal processes on SEM, is well-established [18]. As such, the SEM is a likely source of benchmarking metrics for the efficacy and ecological validity of virtual environments in therapeutic and training applications. Unfortunately, the long tradition of use of the SEM paradigm in basic research in psychophysiology has resulted in a de facto standard of response quantification procedures [19]. This "received approach" may not be amenable to protocols that will be useful for eyeblink-actuated human computer interaction systems.

1.2 Introduction to Current Study

The eye-blink EMG signals used for analysis in this paper were obtained in a study which attempted the SEM paradigm in an investigation of immersion to detect differences in arousal if any, between persons placed in "low" and "high" immersion scenarios. Subjects were asked to passively view a virtual environment (VE) on two separate experimental runs consisting of both a highly immersive (HI) viewing condition and a low immersion (LI) viewing condition. In the HI condition, subjects wore a head mounted display (HMD) with full tracking capabilities and were free to explore their environment visually. The HI condition also made use of headphones and a tactile transducer floor to simulate riding in a large vehicle. The LI condition consisted of watching the same virtual Iraqi scenario on a 17 inch laptop screen while wearing headphones. During the LI condition, subjects viewed the VE from a static position.

The VE was comprised of a series of safe and combat zones in an Iraqi city. In both the HI and LI conditions, subjects viewed the VE from the perspective of the driver of a Humvee. The speed and trajectory of the vehicle was kept constant to control for time spent in each zone of the VE. Safe zones consisted mainly of a road surrounded by a desert landscape and were free of gunfire and other loud noises. The combat zones included improvised explosive devices (IEDs), gunfire, insurgents, and screaming voices. Subjects passed through 3 safe and 3 combat zones on each experimental run. The total length of each run was 210 seconds.

An acoustic startle probe was used to elicit startle eyeblink responses. The startle probe was a 110 dB white noise burst 50 ms in duration with a near instantaneous rise/fall time presented binaurally through Telephonics TDH-50P headphones. Decibel levels were measured with a Realistic sound level meter using a Quest Electronics earphone coupler. Startle probes were experienced intermittently throughout the experimental runs. A total of 4 startle probes were experienced in both the safe and combat zones in each run.

1.3 Overview of Considerations for Response Quantification

According to the "scoring protocol" found within the "received approach" to psychophysiological experimentation, a set of signal parameters of interest and associated procedures to extract them from raw signal recordings must yield outputs that are readily amenable to further standardized analyses and expert interpretations. Due to the long tradition of use of the SEM paradigm in basic research in psychophysiology, the "received approach" represents a de facto standard of response quantification procedures [19]. The translation of psychophysiology procedures from "assessment" of persons' responses to computer mediated information to "adaptive human computer interfaces" presents the requirement of devising scoring protocols for situations to which the existing de facto standards (i.e. received approach) may not be very suitable. It is reasonable to expect that the scoring protocols that will be useful for eyeblink-actuated systems (e.g. BCIs), may be dissimilar to the parameters currently of interest in psychophysiological assessment.

In this paper, some general considerations for designing response quantification procedures in non-traditional psychophysiological applications are presented, using eyeblink scoring for a Humvee Immersion Experiment as a case study.

1.4 Open Parameters in the Design of Scoring Protocols for Electromyographic Eyeblink Recordings

Signal acquisition consoles currently popular for electromyographic (i.e. EMG) signal acquisition [20] typically incorporate pre-processing procedures like amplification. Further signal conditioning stages like filtering and thresholding, which characterize a response quantification procedure, were earlier performed using analog hardware implementations which are now largely being replaced by digital software implementations. For the study reported herein, filtering and thresholding were implemented using MATLAB®. The following are potential open parameters from which the investigator may chose while devising a scoring protocol for raw EMG recordings:

Filter parameters. Following the "received approach", the smoothening (low-pass filtering) requirements are determined largely by the quality of the measuring equipment and likely sources of noise. While smoothing is a necessity for noisy signals, it has the undesirable side-effect of causing phase distortions and in general, a trade-off is sought to give a high signal-to-noise ratio with minimal phase distortion. Throughout this paper, "box-car" filtering (i.e. running average filter) is implemented with the filter width (i.e. window size) as the open parameter. Box-cart filtering was chosen herein for illustrative purposes due to its ease of implementation, computational speed, and near-ubiquitous use in EMG signal conditioning.

Threshold. Following the "received approach", portions of the filtered and rectified eyeblink trace within a small time-window adjacent to a startle-eliciting stimulus are treated as startle blinks if their magnitude exceeds a stipulated threshold. Threshold choice is essentially dictated by typical eyeblink EMG amplitudes observed in the population under study, and must take into account the amplitude reduction caused by the filter settings in the smoothening stage.

Region of interest following the startle blink. Following the "received approach", the length of the time window following a startle-eliciting stimulus within which a super-threshold signal will be considered a blink associated with that stimulus. This is decided mainly on psychophysiological considerations like the average latency of the population under study. Herein the length of this time window is fixed at 150 ms which is psychophysiologically established as a typical time-to-peak for startle-elicited eyeblink signals.

In what follows, results obtained with divergent filter parameter and threshold settings are compared with those obtained from a "reference setting" conforming to the "received approach" to in psychophysiological assessment. This is performed to gain an increased understanding of the effect of these parameter changes and ultimately guide design of augmented scoring protocols for future applications.

1.5 Immediate Motivations for the Present Study

In preliminary analyses, two different scoring methods were considered; 1) Received Approach: the output of this initial protocol being the superthreshold EMG of the smoothed signal in a 150 ms window adjacent to the startle probe; with the threshold at 0.0001 mV; and 2) Elementary Signal-Processing Approach (distortion-causing signal-smoothening is preferably avoided): the score for this alternative protocol being the mean amplitude of superthreshold EMG spikes of the unsmoothed signal in a 150 ms window adjacent to the startle probe, with the threshold at 0.15 mV. An analysis of variance between the EMG scores for "high" and "low" immersion did not report significant differences ($F = 0.091$, $p = 0.763$) for the scoring procedure comporting to the "received approach". Contrariwise, the Elementary Signal-Processing Approach scoring procedure resulted in significant differences ($F = 17.412$, $p < 0.001$). This obvious mismatch between the results of two scoring protocols motivated the investigations presented in this paper to determine feasible ranges of parameter settings for eye-blink scoring protocols. The preliminary findings underscore the need for caution in the choice of scoring metrics and proper validation of the same. These considerations may be of importance to researchers interested in eyeblink analyses to be reliably employed in non-traditional applications psychophysiological data.

2 Methods

Eyeblink EMG traces from the physiological recordings obtained from the orbicularis oculi muscle of subjects in the Humvee Immersion Experiment using Biopac MP150 were exported to MATLAB® and scored using multiple protocols to enable comparison. Each EMG trace was of duration 210 seconds, during which subjects experiencing the Humvee scenario were administered 9 acoustic startle probes via headphones at pre-determined times. The traces were smoothed using a box-car (running average) filter and rectified. Figure (2 a) shows part of a typical EMG trace filtered with running average filters of different filter-widths and rectified. Trace segments of 150 milliseconds adjacent to each startle probe were then examined for spikes whose magnitude exceeded a pre-determined threshold. These super-threshold spikes were treated as startle blinks and a time series of startle-blinks was obtained from each

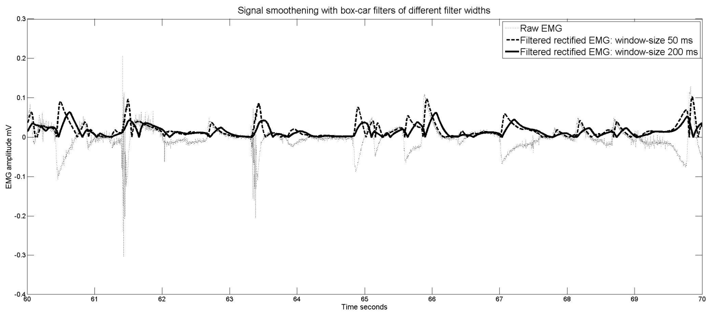


Fig. 1a. Rectification and smoothening of the raw EMG signal (dotted trace) using a running-average filter with window sizes 50 ms (dashed trace) and 200 ms (bold trace)

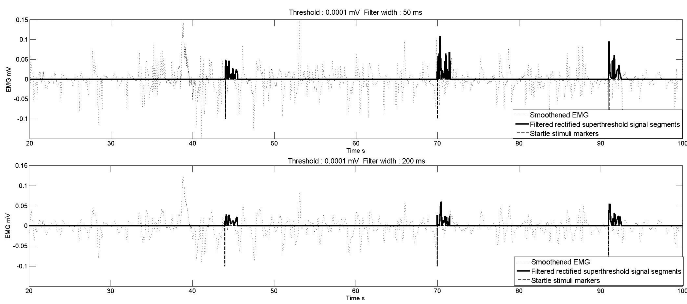


Fig. 1b. Startle blink time series (bold traces) obtained with two different settings of filter window-size and threshold, from the same raw eyeblink signal. The traces shown above are a portion of a 210-second-long electromyographic recording in the Humvee Immersion Experiment with acoustic startle probes used to elicit the blink responses. The positions of these startle probes are shown by the dashed markers.

EMG trace. Figure (2b) shows startle-blinks obtained using this procedure, for different parameter settings. Each parameter setting corresponds to a scoring protocol.

The reference protocol adopted to represent the “received approach”, used a box-car filter window size of 100 samples (corresponding to 100 ms for a sample rate of 1 kHz) and a threshold of 0.1 microvolt, reasonably conforming to current psycho-physiological practice. Alternative protocols were generated by varying the filter window-size from 10 to 200 samples and the threshold from 0.01 microvolt to 5 microvolt. The covariance of the output signals from each of these alternative protocols, with the corresponding output of the reference protocol is treated as an index to study the effect of parameter variations. The covariance between the startle trains obtained from the reference protocol and a test protocol, is of interest because it is a measure of how much the two signals ‘vary together’ i.e. how much a signal tends to be above its expected value if the other signal is known to be above its own expected value. A high positive covariance between the output signal of a particular test protocol and that of the reference protocol would suggest that there is a high degree of linear dependence between the two signals. An implication of such a high degree of linear independence is that the investigator can treat such a test protocol as practically equivalent to the reference protocol.

The results presented in Section 3 represent the mode of the observations for different subjects. The conclusions of this analysis are expected to guide the design of a scoring protocol suited to the ‘most typically encountered’ subject. Therefore the measure chosen for this analysis was the ‘mode’ which represents the most frequently occurring value in the sample.

3 Results

Figure 3 summarizes the effect of two open parameters of interest, namely, boxcar filter width and EMG threshold, on the covariance matching score of test protocol outputs with the reference protocol. The electromyographic recordings used for this analysis were obtained from 7 subject volunteers of the Humvee Immersion Experiment in the high immersion condition. Each curve in the family of curves in the upper panel, shows the effect of filter width on the covariance matching score, for a fixed value of threshold (Upper curves are for lower thresholds.). Likewise, each curve in the lower panel shows the effect of EMG threshold on the covariance matching score, parametric on the filter width. Taken together, these results represent a scan of the space of open parameters to determine feasible settings. A visual examination of the curves reveals that there is ‘plateau’ of very slow decline around the optimal values of maximum covariance outside which the covariance score shows a more rapid decline. The range of values in the ‘plateau’ regions represent parameter settings which provide outputs nearly equivalent to those of the reference protocol. This suggests that traditional parameter settings are not tight and strict, but belong to a larger range of values of open parameters which a designer can explore, while still conforming reasonably to established psychophysiological practice.

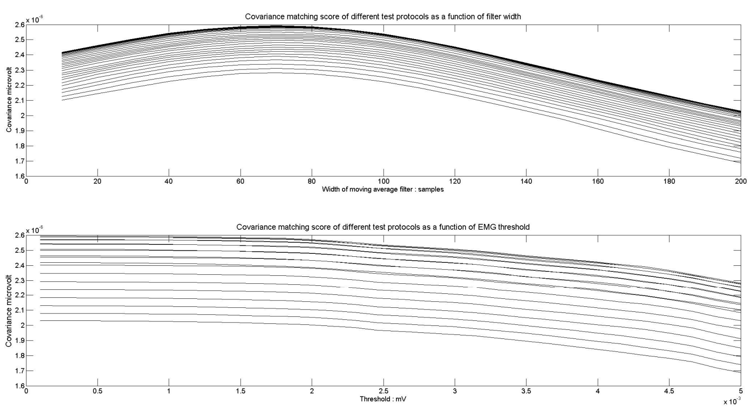


Fig. 2. Effect of filter width (upper panel) and effect of EMG threshold (lower panel) on the covariance matching score of a test scoring protocol with respect to the reference protocol. View text in Section 3 for details and explanation.

4 Discussion

4.1 Consensus between Requirements of Basic Researchers and Application Developers

The “received approach” to scoring protocols parameter settings need not always be viewed as a tight requirement. Instead, as the examples in this paper have illustrated, they may be viewed as a flexible constraint from which a designer may make judicious use of adjustable parameters based on other application specific demands. For example, in an experimental scenario where signal-to-noise ratio is a high priority, the designer can choose the highest filter width within the “plateau” region, thus representing a consensus of the requirements of traditional psychophysiology as well as engineering considerations. Although the method of analysis presented herein is merely suggestive, it can be readily adapted to other scenarios. Further, its variants may be employed to find feasible ranges of parameter settings. As research with psychophysiological modalities becomes more and more interdisciplinary, and the composition of investigating teams becomes more and more diverse, such analyses will be a useful way to arrive at consensus protocols that address the concerns of career psychophysiologists, signal-processing engineers, and designers of virtual environments.

A case in point, of the competing concerns of basic researchers in psychophysiology and signal processing engineers is the need for signal smoothening. Traditional response quantification procedures that are well-established in the psychophysiology research community typically involve signal smoothing of the EMG signal as a pre-processing step, and extraction of peak amplitude and response latency from the smoothed signal. Researchers interested in human computer interfaces (i.e. BCIs), however, may prefer to average raw unsmoothed signals across trials, to avoid the signal distortion and information loss caused by smoothing. A method such as that described above can be useful to arrive at a mutually acceptable parameter setting.

4.2 Emerging Trends and Relevance of Interdisciplinary Dialogue on Response Quantification Standards

In psychophysiological assessment, data scoring is typically performed with ‘off-the-shelf’ scoring software [21], most often proprietary and provided by the manufacturers of the recording equipment. In recent years, the recognition of several new parameters of interest in EMG analysis [22] motivates the need for newer sophisticated data scoring procedures capable of dealing with a greater variety of metrics. Investigators, especially in nascent applications, will need to choose parameters tailored to the application, exploiting the additional capabilities of state-of-the-art data acquisition equipment and using flexible customizable data-scoring software, typically written in MATLAB ®. These developments will lead to a shift from an “off-the-shelf” approach to a “drawing-board” approach in designing response quantification procedures and scoring protocol design. Attempts to contribute (such as the one found in this paper) to the ensuing interdisciplinary discourse to arrive at universally accepted standards for psychophysiological response quantification, deserve due attention so that the full potential of the confluence between psychophysiology and virtual environments is attained in a genuine, well-founded manner.

References

1. Ferreira, A., Celeste, W.C., Cheein, F.A., Bastos-Filho, T.F., Carelli, R.: Human-Machine Interfaces Based on EMG and EEG Applied to Robotic Systems. J. of NeuroEngineering and Rehabilitation, 5–10 (2008)
2. Allanson, J., Fairclough, S.H.: A research agenda for physiological computing. Interacting With Computers 16, 857–878 (2004)
3. Picard, R.W.: Affective Computing. MIT Press, Cambridge (1997)
4. Parsons, T.D., Rizzo, A.A.: Affective Outcomes of Virtual Reality Exposure Therapy for Anxiety and Specific Phobias: A Meta-Analysis. Journal of Behavior Therapy and Experimental Psychiatry 39, 250–261 (2008)
5. Parsons, T.D., Bowerly, T., Buckwalter, J.G., Rizzo, A.A.: A controlled clinical comparison of attention performance in children with ADHD in a virtual reality classroom compared to standard neuropsychological methods. Child Neuropsychology 13, 363–381 (2007)
6. Parsons, T.D., Rizzo, A.A.: Neuropsychological Assessment of Attentional Processing using Virtual Reality. Annual Review of CyberTherapy and Telemedicine 6, 23–28 (2008)
7. Parsons, T.D., Rizzo, A.A., Bamattre, J., Brennan, J.: Virtual Reality Cognitive Performance Assessment Test. Annual Review of CyberTherapy and Telemedicine 5, 163–171 (2007)
8. Parsons, T.D., Rizzo, A.A.: Initial Validation of a Virtual Environment for Assessment of Memory Functioning: Virtual Reality Cognitive Performance Assessment Test. Cyberpsychology and Behavior 11, 17–25 (2008)
9. Parsons, T.D., Silva, T.M., Pair, J., Rizzo, A.A.: A Virtual Environment for Assessment of Neurocognitive Functioning: Virtual Reality Cognitive Performance Assessment Test. Studies in Health Technology and Informatics 132, 351–356 (2008)
10. Parsons, T.D., Larson, P., Kratz, K., Thiebaux, M., Bluestein, B., Buckwalter, J.G., Rizzo, A.A.: Sex differences in mental rotation and spatial rotation in a virtual environment. Neuropsychologia 42, 555–562 (2004)
11. Parsons, T.D., Rizzo, A.A., Buckwalter, J.G.: Backpropagation and regression: comparative utility for neuropsychologists. Journal of Clinical and Experimental Neuropsychology 26, 95–104 (2004)
12. Parsons, T.D., Rizzo, A.A., van der Zaag, C., McGee, J.S., Buckwalter, J.G.: Gender and cognitive performance: a test of the common cause hypothesis. Aging, Neuropsychology, and Cognition 12, 78–88 (2005)
13. Macedonio, M., Parsons, T.D., Rizzo, A.A.: Immersiveness and Physiological Arousal within Panoramic Video-based Virtual Reality. Cyberpsychology and Behavior 10, 508–516 (2007)
14. Parsons, T.D., Iyer, A., Cosand, L., Courtney, C., Rizzo, A.A.: Neurocognitive and Psychophysiological Analysis of Human Performance within Virtual Reality Environments. Studies in Health Technology and Informatics 142, 247–252 (2009)
15. Pugnetti, L., Meehan, M., Mendozzi, L.: Psychophysiological correlates of virtual reality: a review. Presence 10, 384–400 (2001)
16. Grillon, C., Baas, J.M., Cornwell, B., Johnson, L.: Context Conditioning and Behavioral Avoidance in a Virtual Reality Environment: Effect of Predictability. Biol. Psychiatry 60, 752–759 (2006)
17. Baas, J.M., Nugent, M., Lissek, S., Pine, D.S., Grillon, C.: Fear Conditioning in Virtual Reality Contexts: A New Tool for the Study of Anxiety. Biol. Psychiatry 55, 1056–1060 (2004)

18. Harbin, T.J., Berg, W.K.: The Effects of Age and Attention Upon Reflex Inhibition. Biol. Psychol. 22, 81–94 (1986)
19. Blumenthal, T.D., Cuthbert, B.N., Filion, D.L., Hackley, S., Lipp, O.V., van Boxtel, A.: Committee report: Guidelines for human startle eyeblink electromyographic studies. Psychophysiology 42, 1–15 (2005)
20. Carmel, S., Macy, A.J.: Physiological Signal Processing for Biomedical Engineering Education, BIOPAC Systems, Inc., 42 Aero Camino, Goleta, CA, USA, http://www.biopac.com
21. Balaban, M.T., Losito, B., Simons, R.F., Graham, F.K.: Offline Latency and Amplitude Scoring of the Human Reflex Eyeblink with FORTRAN IV. Psychophysiology 23, 612 (1986)
22. Blumenthal, T.D., Melkonian, D.S.: A Model Based Approach to Quantitative Analysis of Eyeblink EMG Response Components. J. of Psychophysiology 17, 1–11 (2003)

Neurophysiological Measures of Brain Activity: Going from the Scalp to the Brain

Phan Luu, Catherine Poulsen, and Don M. Tucker

Electrical Geodesics, Inc., Eugene, Oregon
pluu@egi.com

Abstract. Behavior, such as reaction time and correctness of a response, is the most studied output of the mind in the fields of psychology and human factors. With the advent of modern neuroimaging technologies, opportunities exist for direct study of the mind's machinery: the brain. Moreover, there are opportunities for applying these technologies to solve a host of educational and engineering challenges, such as how to design better interfaces with computer systems or how to better educate and train students. The electroencephalogram (EEG) is a direct reflection of the functioning brain, and technologies that enable recording of the EEG have been in existence for more than 50 years. Within the past decade substantial progress has been made in EEG technology, permitting a direct view into the brain. We cover these advances in this paper, which include dense-sensor array technology and physics-based computational head models, and present several examples of how they have been applied.

Keywords: Dense-Array EEG, neuroergonomic.

1 Introduction

Historically, psychological studies relied on behavioral and self-report data to understand the workings of the mind. With carefully designed studies, researchers were able to explicate various cognitive processes and how these processes may be understood using different models of cognition (e.g., [1]). However, even with carefully designed studies, inferences about cognition remain limited because behavior is the final output of the mind, with many cognitive processes contributing to it. Ultimately, to understand behavior and the human mind, the study of cognitive functions must be placed within the context of brain function. Prior to the advent of modern neuroimaging technologies, relating cognitive processes to brain function was accomplished mainly through animal studies, which permit invasive and destructive techniques, or the study of patients with existing brain lesions. These types of studies also have inherent limitations (such as disruption of brain network dynamics).

With the advent of functional magnetic resonance imaging (fMRI), it became possible to study the human brain non-invasively. The development of the field of cognitive neuroscience quickly followed. Using these new technologies, researchers have confirmed old findings as well as discovered astonishing new ones (e.g., how quickly the brain can reorganize during learning). The field of human factors engineering is now beginning to consider the advantages of understanding human behaviors within

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 488–494, 2009.

the brain framework. This is embodied in the emerging field of neuroergonomics. However, requirements of neuroergonomic studies are different from requirements of cognitive neuroscience studies. For the latter, it is adequate to conduct studies in controlled laboratory settings with specially built facilities. However, the goals of neuroergonomic studies are more practical, requiring studies to be performed in more naturalistic settings. For these studies, fMRI technology is not practical; more portable neuroimaging technologies are required.

Functional near infrared (fNIR) is akin to fMRI; it is sensitive to blood-oxygen level changes and is believed to provide good spatial resolution for the detection of superficial (several centimeters deep) sources of brain activation. These systems are often small and portable, making them suitable for in-field neuroergonomic studies. However, fNIR is insensitive to deeper sources due to the attenuation, through absorption and scattering, of the number of photons at greater depths. Moreover, because photons are also absorbed by hair (particularly dark hair), measurements over most of the head of typical subjects are problematic.

The EEG, like fNIR, is a portable technology that can be made field deployable. Although it is true that the electroencephalogram (EEG) existed long before fMRI and that it can be considered a neuroimaging technology, it was used mainly as a technique that supplemented behavioral analyses. That is, researchers often treated the EEG (or its derivatives, such as the event related potential – ERP) as a dependent variable akin to reaction times in behavioral studies, very rarely referring to or considering the brain that generated the signals. This may be attributable to the fact that only a few sensors were ever used in these early studies, making it difficult to relate the EEG to the underlying neural generator. With recent advances, however, it is now possible to relate the EEG to specific activations of the underlying cortex, and thus bring the full power of whole-head, direct neuroimaging to neuroergonomic applications.

For EEG to be useful as a neuroimaging technology several requirements must be satisfied. First, the scalp electrical field, which is recorded as the EEG, must be accurately described because it is from this description that estimates of the underlying cortical sources are derived. Inaccuracies in this description result in inaccurate estimates. Because current must flow through resistive tissue (such as the skull), a point-to-point correspondence between the location of the electrode on the scalp and an activated cortical patch cannot be assumed. Therefore, a second requirement is the development of head models that accurately describe the propagation of current from the cortical sources to scalp electrodes (i.e., the lead field). Once these two requirements are met, various inverse methods can be employed for source estimation [2]. In the following sections we briefly describe these two requirements.

Accurate description of the scalp potential field: Spatial resolution is simply dependent on the density of spatial sampling. The Nyquist theorem for discrete sampling is well appreciated by electroencephalographers in the time domain. When the time series of an EEG channel is sampled discretely, the highest measurable frequency is half the sampling rate. If the signal contains information at higher frequencies than the Nyquist limit, they are not only poorly characterized: they alias or appear misleadingly as increased energy at lower frequencies. If the EEG time series has been aliased because of undersampling, there is no valid method for removing or undoing the distortion by digital signal processing methods. The Nyquist criterion also applies to spatial sampling [3].

The scalp surface potential at any point in time is a continuous field over the surface of the head. The sensor (electrode) array effects a discrete sampling of this field, and like any discrete sampling is subject to a Nyquist criterion. However, unlike the EEG time series, the spatial signal is acquired discretely. While the time series at each channel can be analog low-pass filtered prior to digitization, the spatial signal has information content fixed by both the source distribution and the number of electrodes. As a consequence, any aliasing due to undersampling cannot be undone, and it is critical that an adequate sampling of the potentials be accomplished from the outset. Assuming an even spacing of electrodes, the inter-electrode distance determines the highest spatial frequency that can be observed without aliasing. In addition to the requirement of inter-sensor distance, adequate spatial sampling also requires that the inferior surface of the head be captured, particularly if one wishes to capture the activity of the basal surface of the brain, including the orbitofrontal and basal temporal lobes.

Accurate head model: Accurate head models require accurate description of 1) the geometry of each head tissue, 2) the conductivity values of each head tissue, 3) the location and orientation of potential sources, and 4) the position of the electrode sensors on the scalp. Together, these four pieces of data form what is referred to as the forward model (or lead field) used for source modeling of scalp-recorded data.

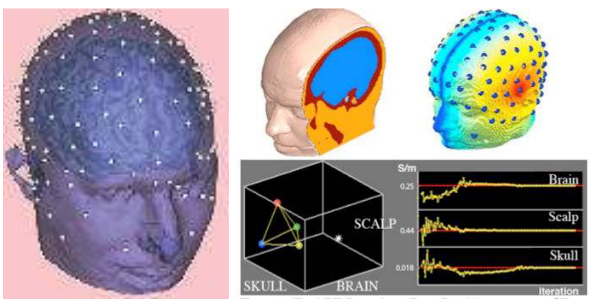


Fig. 1. Left: Head model with sensors registered to scalp surface. Right: 3D subject geometry, CT registered with MRI, the forward finite difference method solution for a particular set of tissue conductivities. Sensor positions, acquired with the Geodesic Photogrammetry System, are registered with the head model (top right).

With dense-array sensors and accurate head models, we can compare source estimate results with fMRI findings to determine convergence. In this paper, we start with a brief review of a dual-system learning model derived from a meta-analysis of fMRI results. This meta-analytic study identifies cortical structures involved in the learning process. We then qualitatively compare these findings with results from two learning studies conducted with dense-array EEG [4, 5].

2 Method

In a meta-analytic review of human learning and performance studies using fMRI technology, Chein & Schneider [6] identified two systems that are differentially engaged during learning and performance. They showed that the prefrontal lobes (including the inferior frontal gyrus, dorsolateral prefrontal cortex, and medial prefrontal cortex) and anterior cingulate cortex (ACC) are strongly engaged early in the learning cycle when stimulus-response mappings are actively being established. In later stages of learning, however, once contingency mappings have become consolidated, these frontal structures exhibit a reduction in activity. In contrast, posterior regions, including the posterior cingulate cortex (PCC), precuneus, cuneus, superior parietal lobule, and intraparietal sulcus demonstrate increased activity during these later stages. The reduction of activity observed in frontal structures in later stages of learning appears to represent reduced reliance on top-down executive control systems once learning has been established, while the increased activity observed in posterior structures may represent the establishment and automatization of the learned action patterns as well as continued monitoring of performance. Figure 2 presents the location of some of these areas. These data will serve as the comparison for the dense-array EEG source estimate results.

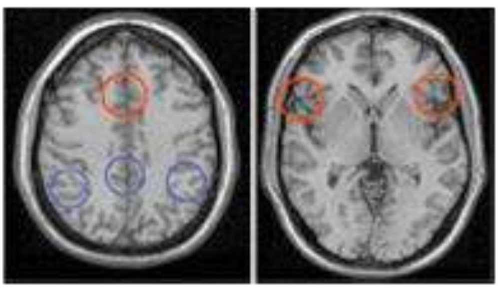


Fig. 2. Cortical structures that make up the dual-system model. Red circles mark cortical regions that make up the early-learning system. These include the ACC and inferior frontal gyrus (shown). Blue circles mark the cortical regions that make up the late-learning system. These include the PCC and the superior parietal lobules (shown). See text for other components.

The dense-array EEG data are taken from two learning studies we conducted previously [4, 5]. In these studies, participants were required to learn stimulus-response mappings through trial and error. That is, on some trials, participants had to determine whether a two digit code (e.g., 12) required a response or not, and if it required a response, they head to determine whether to respond with the index or middle finger of either the right or left hand. After each response, participants were provided with a feedback stimulus to inform them of the correctness of their response. All participants used the feedback to guide their learning. Participants saw a total of 16 two-digit numbers (8 required a response and 8 required no response).

We used a 'fixed-number of consecutive correct responses' method for determining a learning threshold for each participant. In this method, the learning threshold is defined as the moment when participants made four consecutive correct responses (or withholding of a response) for a particular stimulus. On average, participants took 13 trials with each stimulus to learn the stimulus-response mapping.

The EEG data were sorted according to pre-learned and post-learned trials. These trials types were averaged to form an ERP. The grand-average of all participants' ERPs was then submitted to source estimation using a lead field that was constructed with a finite difference model (FDM). The FDM allows accurate characterization of the cranial orifices, primarily the optical canals and foramen magnum. Tissue compartments of the FDM were constructed from whole-head MRI (Colin_27) and CT scans of a single subject whose head shape closely matches the Montreal Neurological Institute (MNI) average MRI (MNI305). The MRI and CT images were co-registered prior to segmentation of the brain and cerebral spinal fluid (identified from MRI data), and the skull and scalp (identified from CT images), and the individual's MRI and CT images were aligned with the cortex volume from the MNI atlas with Talairach registration. The tissue volumes were parceled using 2 mm voxels to form computational elements of the FDM. Conductivity values used in the FDM model are as follows: .25 S/m (Siemens/meter) for brain, 1.8 S/m for cerebral spinal fluid, .018 S/m for skull, and .44 S/m for scalp. These values reflect recent evidence that the skull-to-brain conductivity ratio is about 1:14, compared to the 1:80 ratio traditionally assumed.

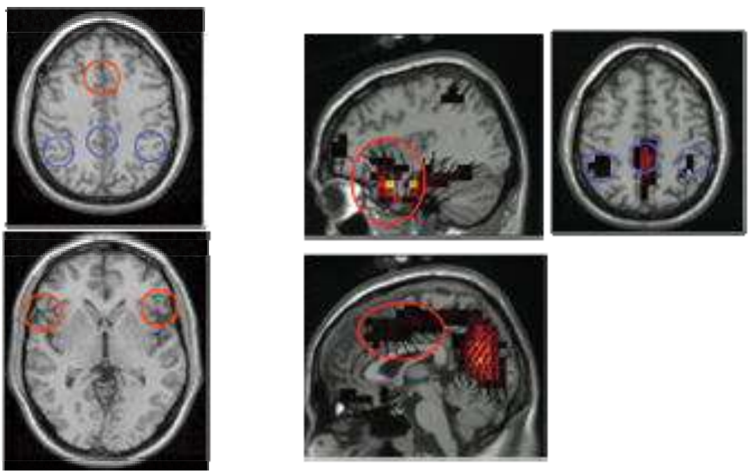


Fig. 3. Left: See Figure 2 caption. Right: Activity localized to the inferior frontal gyrus and ACC (early-learning system) and PCC and parietal lobules (late-learning system).

3 Results

Figure 3 shows the source estimate results from the two learning studies [4, 5] compared to the dual-system learning model. Of particular significance was that, as predicted, we found activity in the inferior frontal gyrus, ACC, PCC, parietal lobules, cuneus, and precuneus. We found that the inferior frontal gyrus and medial frontal lobes were indeed involved in the early stages of learning, as predicted, but that their activity only decreased after learning in response to viewing the feedback stimuli. When viewing the target (i.e., the imperative stimulus), activity in the inferior frontal gyrus and ACC remained strong, even after learning. We also found that activity localized to the PCC, parietal lobules, cuneus, and precuneus increased after learning, as predicted by the dual-system model.

4 Discussion

There are existing studies that have simultaneously recorded EEG and fMRI data. However, these studies often employed sparse-array channel counts (e.g., [7]) and, therefore, EEG source estimate results were poor. Nevertheless, these studies demonstrated good convergence between EEG and fMRI data. The results that we obtained from our dense-array EEG studies during learning confirm the predictions made by the dual-system learning model. This model predicts that there are separable early- and late-learning systems that contribute uniquely to the learning process. This learning model was constructed by Chein and Schneider [6] based on data from human fMRI studies. This model is also supported by data from animal studies that point to the existence of similar systems. Therefore, the convergence we found between our source results with the predictive model is quite striking.

However, unlike fMRI and animal studies, data from dense-array EEG studies provide exquisite temporal information because the EEG directly reflects neuronal activity.. We found that activities localized to structures of the early-learning system occur quite rapidly (~ 250 ms after onset of the imperative stimulus) and that they substantially overlap in time with activities localized to structures of the late-learning system (~ 400 ms after onset of the imperative stimulus). The temporal course of these activations reflects the dynamic nature of neurophysiological function.

In summary, evidence shows that source estimates derived from dense-array EEG data may have the spatial resolution required for studying brain processes involved in cognition. Comparison of EEG results with other neuroimaging modalities, such as fMRI, is now possible because EEG data can be analyzed within the same space (i.e., neural source space rather than scalp space) as the analysis of fMRI data. As more EEG studies demonstrate such convergence of results, the utility and validity of dense-array EEG for neuroergonomic studies will be realized.

References

1. Posner, M.I.: Chronometric explorations of mind. Lawrence Earlbaum Associates (1978)
2. Koles, Z.J.: Trends in EEG source localization. Electroencephalography and clinical Neurophysiology, 127–137 (1998)
3. Srinivasan, R., Tucker, D.M., Murias, M.: Estimating the Spatial Nyquist of the Human EEG. Behav. Res. Meth. Inst., Comput. 30, 8–19 (1998)
4. Luu, P., Shane, M., Pratt, N.L., Tucker, D.M.: Corticolimbic Mechanisms in the Control of Trial and Error Learning. Brain Res. 1247, 100–113 (2009)

5. Luu, P., Tucker, D.M., Stripling, R.: Neural Mechanisms for Learning Actions in Context. Brain Res. 1179, 89–105 (2007)
6. Chein, J.M., Schneider, W.: Neuroimaging studies of practice-related change: fMRI and meta-analytic evidence of a domain general control network for learning. Cog. Brain Res. 25, 607–623 (2005)
7. Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., Yves von Cramon, D., Engel, A.K.: Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamic of performance monitoring. J. Neurosci. 25, 11730–11737 (2005)

Parsimonious Identification of Physiological Indices for Monitoring Cognitive Fatigue

Lance J. Myers and J. Hunter Downs

Archinoetics, LLC. 700 Bishop St, Suite E, Honolulu, HI, 96813

Abstract. The objective of this study was to identify a parsimonious set of physiological measures that could be used to best predict cognitive fatigue levels. A 37 hour sleep deprivation study was conducted to induce reduced levels of alertness and cognitive impairment as measured by a psychomotor vigilance test. Non-invasive, wearable and ambulatory sensors were used to acquire cardio-respiratory and motion data during the sleep deprivation. Subsequently 23 potential predictors were derived from the raw sensor data. The least absolute shrinkage and selection operator, along with a cross validation strategy was used to create a sparse model and identify a minimum predictor subset that provided the best prediction accuracy. Final predictor selection was found to vary with task and context. Depending on context selected predictors indicated elevated levels of sympathetic nervous system activity, increased restlessness during engaging tasks and increased cardio-respiratory synchronization with increasing cognitive fatigue.

Keywords: cognitive fatigue, heart rate variability, feature selection, wearable sensors.

1 Introduction

Fatigue is a growing problem in modern society. Although sleep experts have found that most adults need 8 hours of sleep per night [1], the average American adult is sleeping only 6.8 hours per night [2], and as much as 20% of the population appears to be acquiring only 6.5 hours of sleep per night [3].

In general terms, losing even small amounts of sleep each night will exert cumulative adverse effects on waking performance which include vigilance decrements, increased lapses of attention, cognitive slowing, short-term memory failures, deficits in frontal lobe functions, and rapid and involuntary sleep onsets [4]. Studies comparing the effects of increased blood alcohol concentrations (BAC) to the effects of sleep loss illustrate the seriousness of insufficient sleep on alertness and performance. Investigations have shown that sustained wakefulness of 20-24 hours produces decrements equal to those observed with BAC levels of between 0.08%-0.10% on tests of psychomotor performance, grammatical reasoning, vigilance, and simulated driving performance [5]. Operator fatigue is frequently responsible for costly accidents and mishaps in driving, aviation, shift health care workers and other similar industries.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 495–503, 2009.

If it were possible to accurately predict when an individual was becoming overly fatigued, timely mitigation strategies could be employed to prevent accidents and other costly fatigue related problems. Several computer models currently calculate performance-readiness predictions that are generally accurate, but none are capable of accounting for individual differences in fatigue vulnerability [6].

Clearly, sleep-wake times are insufficient to model the impact of individual differences in fatigue vulnerability. Therefore we hypothesized that a more accurate assessment of cognitive fatigue could be made by using measured physiology of an individual. To address this, we first ask the question of which physiological variables would be the best predictors of cognitive fatigue. This paper focuses on this question and seeks to identify a parsimonious subset of physiological variables that best track changes in cognitive fatigue and vigilance due to chronic sleep restriction.

2 Methods

2.1 Physiological Measurements

2.1.1 Sensor Technology

There are many candidate physiological measurements that may have strong predictive power of cognitive fatigue. Given that real-world monitoring applications require a product that may be used in non-laboratory, ambulatory contexts, we focused only on those sensor technologies that met criteria of non-invasive, ambulatory, wearable, unobtrusive and artifact resistant. We thus selected a representative commercially available ambulatory, wearable monitoring system called the BioHarness (ZephyrTech, NZ). This system is a chest strap that is capable of non-invasively, continuously and simultaneously monitoring electrocardiography (ECG), respiration, motion, posture and skin temperature. The strap is light-weight, comfortable and uses proprietary smart-fabric sensing technology. Data is continuously logged using onboard flash memory and subsequently downloaded to a PC for further offline processing. Data may also be wirelessly acquired over a BlueTooth connection for further real-time development.

2.1.2 Candidate Feature Derivation

There is a strong link between cognitive fatigue and cortical arousal as measured with electroencephalographic (EEG) activity [7]. However, due to the fact that EEG measurements are better suited for laboratory conditions, we sought to investigate whether systemic autonomic arousal demonstrated similar progression to cortical arousal. Studies have demonstrated that the fatigue state is associated with a shift of sympathovagal balance toward sympathetic predominance and reduced vagal tone [7]. Thus measurements indicative of autonomic balance could have sufficient predictive power to discriminate between differing levels of cognitive fatigue.

Heart rate variability

The analysis of heart rate variability provides a way to non-invasively study the autonomic nervous system (ANS) by acting as a dynamic window into autonomic function and balance. Over the years, a variety of metrics have been proposed to succinctly quantify HRV and the associated respiratory sinus arrhythmia (RSA). Due to the variety of metrics, in 1996 the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology suggested standard mathematical procedures for short-term HRV evaluation [8]. However the task force report only considers HRV from the ECG and a number of laboratory studies have documented that changes in respiratory parameters can seriously confound the association of RSA and cardiac vagal tone [9]. Thus when respiratory measures are also available, several adaptations to these methods, as well as new methods may be included. In addition further indices derived from non-linear dynamical analysis are also included. Table 1 lists all the candidate predictor features derived from HRV and RSA measures.

Table 1. Candidate HRV features

Feature name	*Feature description*
Time domain features	
ANN	Average heart rate
RMSSD	Root mean square of successive differences
SDNN	Standard deviation
Frequency domain features	These were derived using the Welch method.
HF	Spectral power in high frequency range[a]
LF	Spectral power in the low frequency range[b]
VLF	Spectral power in the very low frequency range[c]
LF/HF	Ratio of LF to HF
HFnorm	HF power normalized to total power
LFnorm	LF power normalized to total power
Non-linear features	
SampEn	Sample Entropy [10]
CD	Correlation Dimension [10]
MLE	Maximum Luypanov Exponent [10]
Cardio-resp features	
RSA	Correlation coefficient[d]
PSI	Phase Synchrony Index [11][e]

[a]Default range is 0.15-0.4Hz. However, the respiratory rate was calculated from the respiration band and the HF range was centered at the respiration frequency, with a width of 0.15Hz on either side of this.

[b]Default range is 0.04-0.15Hz

[c]Default range is 0.01-0.04Hz

[d]This was the normalized correlation coefficient between the respiratory signal and the heart rate signal over the specified time epoch. The HR signal was resampled using a linear interpolation to an even rate of 2Hz. The respiration signal was downsampled to 2Hz.

[e]This is a newly developed data analysis technique based on the mathematics of nonlinear dynamics and allows any interaction that does occur in even weakly coupled complex systems to be observed [11]. It is especially well suited to probe the weak interactions between irregular and non-stationary oscillators such as the human heart and respiratory system. Phase locking of respiratory and the cardiac rhythms,

Table 2. Candidate respiratory features

Feature name	*Feature description*
General respiration	
RR	Respiration rate
Vol	Tidal volume (uncalibrated)
Mvol	Minute Ventilation
DC	Duty Cycle
Respiratory irregularity	
TVI	Tidal Volume Instability [12]
DCvar	Duty Cycle Variability[12]
BRV	RMS of sucessive differences of breath period [12]

and respiratory modulation of heart rate (RSA), are two competing aspects of cardio-respiratory interaction.

Respiratory features

Everyday observation suggests that psycho-physiologic state is a determinant of respiration and this relationship has been actively investigated dating as early as the beginning of this century. However, contemporary research showing how respiration may be used as a surrogate of autonomic balance is relatively sparse. Despite this, there is evidence that changes in autonomic balance in general do influence respiration, and that different states give rise to different breathing patterns. There are certainly obvious markers such as respiration rate and volume that are typically used to index autonomic changes. However in addition to this several studies point to dysregulated breathing as having the most potential to index arousal. For example, studies examining the relationship between respiration and state in the clinical context of anxiety, panic disorder and chronic pain all point to irregularity in breathing as a key marker of anxiety [12]. Furthermore, it has been demonstrated that irregularity in breathing appears to increase under conditions of sympathetic arousal such as emotional upset and excitement. Table 2 lists all the candidate predictor features derived from respiratory measures.

Motion features

In addition to the cardio-respiratory features, two features derived from the trunk accelerometer were also used. These features were fairly simple features and were the overall level of motion and the variability in motion. These are listed in table 3.

Table 3. Candidate motion features

Feature name	*Feature description*
motion	Mean acceleration
motionVar	Standard deviation of acceleration

2.2 Vigilance Assessment

Objective cognitive performance evaluations were accomplished with the Psychomotor Vigilance Task (PVT). This is a portable, low-voltage, battery-powered reaction-time test known to be sensitive to sleep loss [13]. The PVT is commonly utilized to track changes in vigilance or sustained attention; attributes that underlie the successful accomplishment of many types of more complex cognitive tasks. A variety of data is generated from the PVT, but of primary interest were reaction-time measures, accuracy measures, and attention-lapse indications since these reflect cognitive slowing and response failures.

2.3 Study Design

Six subjects were recruited for a continuous 37 hour sleep deprivation study. Participants were instructed to obtain a minimum of 8 hours of sleep prior to reporting the research facility. They were instructed to awaken at 0700 on day 1 and they subsequently had to remain awake until at least 1900 hours on the second day of the study. Participants were outfitted with the BioHarness and data quality was verified to ensure correct fit and function.

Starting at 1000 on Day 1, each subject had to complete the first 10-minute test session on the PVT. Subsequent PVT sessions occurred every hour until 1800 on Day 2. Before and after each PVT, participants reported to the testing room where they were seated in a comfortable chair and asked to remain relaxed, still, and quiet with eyes open for 3 full minutes to stabilize autonomic activity. In between PVT sessions, participants will be free to play computer games, watch TV, or engage in any other type of sedentary activity. Following completion of the study, data was downloaded off the on-board flash memory storage to a PC for subsequent analysis.

2.4 Parsimonious Feature Selection

The objective of this study was to identify a parsimonious set of physiological measures that could be used to best predict cognitive fatigue levels. The importance of parsimony in feature selection is emphasized as it tends to improve prediction performance and simpler models are preferred for the sake of scientific insight and interpretation of the chosen features. In general, statistical learning theory poses a structural risk minimization criterion that balances the trade-off between good empirical performance (i.e., classification accuracy on training data) and good generalization ability (i.e., classification accuracy on unseen data). Most classifiers will generalize badly in the situation of many irrelevant features. Unfortunately, a frequently encountered constraint when working with physiological data is that the numbers of potential or candidate predictors tend to be of the same order of magnitude or larger than the number of available observations for training. We therefore sought to perform effective feature subset identification given this constraint.

A complete set of candidate features that compactly represent the original physiological data set were identified. A total of $M=23$ features were selected for the candidate feature set. All features were extracted over a 3 minute quiet period preceding and following the PVT test and also the 10 minute period during the PVT test. Each feature was subsequently normalized to the initial baseline period.

For each subject, and for each of time period, a vector of features plus a constant term was formed:

$$\boldsymbol{x}_i = [1, x_{i1}, x_{i2}, \dots, x_{iM}]^T \tag{1}$$

The mean reaction times from the PVT tests were pooled together for each subject and each test to form a response vector:

$$\boldsymbol{y} = [y_1, y_2, \dots, y_N]^T \tag{2}$$

where N is the number of time periods multiplied by the number of subjects. All features were pooled together into a single N by M+1 matrix, $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}, \dots, \boldsymbol{x}_N]^T$.

When the number of predictors M exceeds the number of training samples N, the modeling problem is underdetermined, or ill-posed in the Hadamard sense. In this instance, it is desirable to find a model with significantly fewer predictors and in fact, the more sparse the model, the more likely that the predictors are causally related to the dependent variable. In order to achieve this for ill-posed data, the values of the regression coefficients $\boldsymbol{b}$, can be constrained via a shrinkage function.

$$\boldsymbol{b} = \arg\ \min_b \|\boldsymbol{y} - \boldsymbol{X}b\|^2 + \lambda \sum_{i=1\dots M} |b_i| \tag{3}$$

where λ is a tuning parameter.

This method is called the Least Absolute Shrinkage and Selection Operator (LASSO), and is a regularization and selection method [14]. For each value of the tuning parameter, the LASSO will generate a sparse solution by setting many of the parameters to 0. The LASSO is a particularly attractive algorithm as it uses L^1 norm which can be viewed as the most selective shrinkage function that remains convex. Since a convex function has a global minimum and no local minima, convexity guarantees that we can find the one global solution for a given dataset. We used a highly efficient algorithm for solving the LASSO, termed Least Angle Regression (LARS) [15]. This algorithm converges to the final solution in M steps.

For each solution returned by the LARS algorithm, we evaluated the accuracy of the result using a leave-one-out cross validation procedure. We iterated through each subject, forming a test vector of PVT scores for that subject and a training matrix with the remaining subject. At each iteration we calculated a first order correlation coefficient between the predicted PVT scores and the test PVT scores. These were then averaged to provide a single statistic. The set of features that provided the largest statistic was subsequently selected as the final reduced feature set.

Both predictor and response data were centered and normalized to unit deviation prior to running the feature selection algorithms.

3 Results

Table 4 shows the LASSO selected predictors for each studied epoch. Also displayed are the regression accuracies using the leave-one-out cross validation procedure. No more than two features were selected for each task and this may be due to the fact that many features are well correlated amongst themselves.

As an illustration of prediction performance, figure 1 shows the prediction results for the 3 minute post PVT test period. Note that the data in this figure is normalized and centered to unit deviation.

Table 4. Selected features and prediction accuracy

Task	Selected features	Accuracy
3 minutes prior to test	BRV; LFnorm	0.52
PVT test	Motion	0.74
3 minutes post test	BRV; PSI	0.72

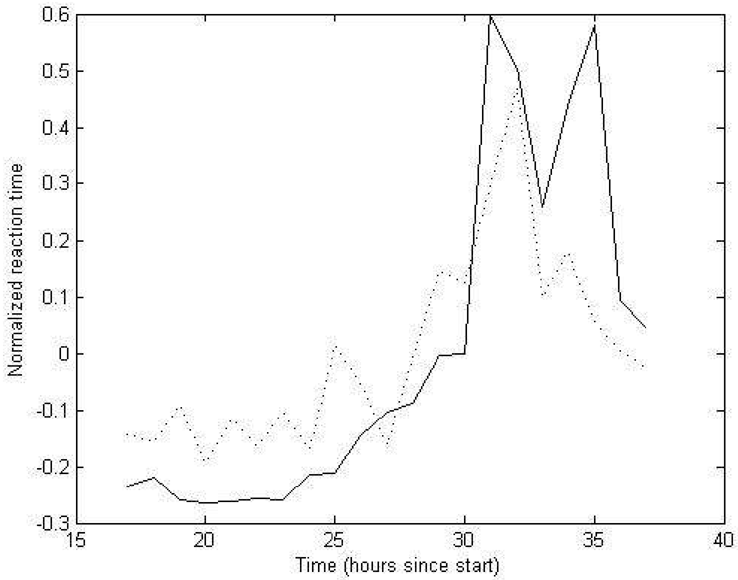


Fig. 1. Prediction results for the 3 minute period post PVT test. The solid line shows the normalized and centered mean PVT reaction time for each continuous hour of wake across subjects. The dotted line shows the regression prediction using the LASSO selected features (BRV; PSI).

4 Discussion

In-between PVT sessions, the participants were free to engage in any type of activity. As a result of this, the levels of autonomic activity during the 3 minute quiet period prior to the PVT test may have been dominated by the activities during the free periods. This is potentially why this period showed the poorest correlations. However, it remains interesting to note that the features selected during this period both strongly represented sympathetic arousal (the breath rate variability and the normalized low frequency power of the HRV). Although it is expected that sleep restriction increases sympathetic system outflow, the weak prediction accuracy may have been due to additional changes in this system caused by uncontrolled activity during the non-test periods.

The feature selection results during the actual PVT test initially appear to be surprising. Here the mean level of body motion was selected as the strongest predictor of cognitive fatigue. Taking a PVT test demands a degree of concentration and during alert periods subjects were more focused and able to concentrate on the task with minimal distraction. However, as the subjects became more fatigued, they tended to become increasingly restless during administration of the PVT test. This restlessness was captured by background body movements as recorded by the accelerometer and was directly correlated with PVT performance.

In the 3 minute stationary rest period following the test, the selected features were the breath rate variability (BRV) and the phase synchrony index (PSI). The BRV represents increases in sympathetic outflow and it has previously been noted that this increases during sustained attention tasks with increasing sleep restriction. The PSI represents the synchronization between the cardiac and respiratory systems and this increase in synchronization only manifested during the rest period following a demanding attention task. It is has been shown that synchronization and modulation (RSA) are two different competing aspects of cardio-respiratory interaction. Often when synchronization goes up, RSA goes down [11] and thus it may indicate changes in sympathetic activity. It is further possible that synchronization plays a homeostatic role in returning the system to baseline levels following increases in sympathetic system outflow. This would explain why there is an increased synchronization drive following high attention tasks when sleep restricted. However further research is required to better understand this feature.

This study demonstrates the power of the LASSO based feature selection paradigm to select a parsimonious physiological feature set. With appropriate individual sleep restriction data, this method could be used to perform individualized feature selection accounting for individual differences in fatigue vulnerability. It is also important to note that accurate context identification is of fundamental importance for any automated fatigue prediction system.

5 Conclusion

The LASSO feature selection technique allows one to select en-masse, via a continuous subset optimization, the set of variables that together are effective predictors of operator alertness status. This technique combined with commercially-available, wearable physiologic monitoring systems is a further step toward a system that can improve operational safety and effectiveness by accurately assessing cognitive fatigue levels during stressful day to day conditions.

References

1. Kleitman, N.: Duration of sleep. In: Sleep and Wakefulness, pp. 114–121. University of Chicago Press, Chicago (1963)
2. National Sleep Foundation. In: 2005 Sleep in America poll. Washington, National Sleep Foundation (2005)
3. Kripke, D.F., Garfinkel, L., Wingard, D.L., Klauber, M.R., Marler, M.R.: Mortality associated with sleep duration and insomnia. Arch Gen Psychiat 59, 131–136 (2002)
4. Carskadon, M.A., Roth, T.: Sleep restriction. In: Monk, T.H. (ed.) Sleep, sleepiness and performance. John Wiley & Sons, New York (1991)

5. Arnedt, J.T., Wilde, J.S., Munt, P.W., Maclean, A.W.: Simulated driving performance following prolonged wakefulness and alcohol consumption: Separate and combined contributions to impairment. J. Sleep Res. 9, 233–241 (2000)
6. Gundel, A., Marsalek, K., ten Thoren, C.A.: Critical review of existing mathematical models for alertness. Somnologie - Schlafforschung und Schlafmedizin 11, 148–156 (2007)
7. Zhang, C., Zheng, C., Yu, X.: Automatic recognition of cognitive fatigue from physiological indices by using wavelet packet transform and kernel learning algorithms. Expert Systems with Applications 36, 4664–4671 (2009)
8. Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology: Heart Rate Variability. Standards of Measurement, Physiological Interpretation, and Clinical Use.: Circulation 93, 1043–1065 (1996)
9. Grossman, P., Kollai, M.: Respiratory sinus arrhythmia, cardiac vagal tone, and respiration: Within- and between-individual relations. Psychophysiology 30, 486–495 (1993)
10. Rajendra Acharya, U., et al.: Heart rate variability: a review. Medical and Biological Engineering and Computing 44, 1031–1051 (2006)
11. Lotric, M.B., Stefanovska, A.: Synchronization and modulation in the human cardiorespiratory system. Physica A: Statistical Mechanics and its Applications 283, 451–461 (2000)
12. Wilhelm, F.H., Trabert, W., Roth, W.T.: Physiologic instability in panic disorder and generalized anxiety disorder. Biological Psychiatry 49, 596–605 (2001)
13. Dinges, D.F., et al.: Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. Sleep 20, 267–277 (1997)
14. Tibshirani, R.: Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological) 58, 267–288 (1996)
15. Efron, B., et al.: Least Angle Regression. The Annals of Statistics 32, 407–451 (2004)

In-Helmet Oxy-hemoglobin Change Detection Using Near-Infrared Sensing

Erin M. Nishimura, Christopher A. Russell, J. Patrick Stautzenberger, Harvey Ku, and J. Hunter Downs III

Archinoetics, LLC, 700 Bishop St, Ste 2000, Honolulu, HI 96813
erin@archinoetics.com

Abstract. Near-infrared (NIR) sensing in flight applications can provide critical objective indicators of crew state. By monitoring oxy-hemoglobin concentrations, a NIR sensor can detect changes in flight crew physiology in response to both cognitive demands and extreme conditions related to flight applications, including gravity-induced loss of consciousness (G-LOC) and hypoxia. A custom NIR sensor was created for in-helmet monitoring of oxy-hemoglobin in flight. This wearable, wireless sensor addresses requirements for flight applications and was applied to a case study that examines the raw optical signal and oxy-hemoglobin response to Valsalva maneuvers performed at 1*g*.

Keywords: Near-infrared sensing, functional brain imaging, oxy-hemoglobin, hemodynamics, physiologic monitoring.

1 Introduction

Physiologic monitoring is currently not widely used in flight applications but can provide critical signals that serve as objective indicators of crew state. An oxy-hemoglobin concentration monitor for in-flight use can respond to both the high cognitive requirements of aircrew as well as the physiologic responses to conditions specific to flight applications. As a measure of hemodynamic response, near-infrared imaging can detect cognitive tasks in a highly-demanding, multi-tasking flight environment for intelligent presentation of subsequent tasks or communications based on current cognitive loads [1]. Monitoring cerebral oxy-hemoglobin concentrations can reflect physiologic changes evoked by high-G maneuvers or hypoxic conditions [2,3]. The effects of such conditions are noteworthy. Between 1982 and 2002, 559 incidents were reported to the USAF Safety Center relating to gravity-induced loss of consciousness (G-LOC), with 30% of these incidents resulting in crashes for single-crewmember aircrafts. The fatality rate for these crashes is 100% [4]. Symptoms of hypoxia range from dizziness, fatigue, and visual impairment to loss of consciousness [5]. These symptoms often go unrecognized by flight crew experiencing them.

To this end, a near-infrared (NIR) sensor has been created for monitoring oxy-hemoglobin changes from within a flight helmet. This wearable sensor was designed to be mounted within a helmet, above the ear cup, and features wireless communications and battery power to eliminate any tethering between the flight crew and the

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 504–513, 2009.

cockpit. The enclosure for the sensor is focused on comfort for long-duration wear, but also allows for adjustability of the optics for proper NIR sensing and protects the electronics from moisture.

The capabilities of this sensor were verified in a case study. Two subjects performed Valsalva maneuvers while wearing the NIR sensor and the oxy-hemoglobin concentrations output from the sensors were analyzed for expected signal changes in response to the maneuvers. The results of this study verify both the functionality and the sensitivity of the system to regional oxy-hemoglobin concentration changes.

1.1 Near-Infrared Sensing

Near-infrared (NIR) sensing technology utilizes the known optical properties of near-infrared light in human tissue to monitor oxy- and deoxy-hemoglobin concentrations [6]. There are two forms of interactions between light and tissue, absorption and scattering, with scattering being the more prevalent form. In between visible and infrared light, NIR light (750nm-2500nm) is relatively weakly absorbed and scattered [7]. Oxy- and deoxy-hemoglobin, present in all living human tissue, have distinct absorption spectra of near-infrared light, which is used to calculate the concentrations of oxy- and deoxy-hemoglobin in tissue based on the differential changes in received NIR light passed non-invasively through this tissue.

By applying this sensing to the head, the oxy-hemoglobin concentration changes can be measured on the cortical surface of the brain. These changes measure the hemodynamic response evoked by cognitive activity in the monitored area, allowing for functional brain imaging. Changes in oxy-hemoglobin concentrations in the head can also be measured as the physiologic response to extreme settings, such as high-G exposures or hypoxic conditions.

Although fMRI is a technology that measures similar physiologic signals related to cognitive activity, the hardware required for near-infrared sensing can be much more compact and less costly. While not offering the same spatial resolution as fMRI, functional near-infrared imaging (fNIR) is more suitable for mobile and/or real-world applications in natural environments.

1.2 In-Helmet Sensor Design

Archinoetics created a custom near-infrared sensor for oxy-hemoglobin change monitoring. The targeted application was for flight crew so the design focused on sensor miniaturization for mounting within a helmet without compromising any safety features. Wireless operation, both for power and communications, was also a priority to prevent tethering between the user and the cockpit, as is preferred for flight platform integration.

The wearable near-infrared sensor created, shown below in Figure 1, features non-invasive NIR sensing through hair for oxy- and deoxy-hemoglobin concentration monitoring with wireless data communications and battery power within a sweat-proof enclosure. In addition to the real-time, continuous-wave NIR signals, the sensor also provides 3-axis accelerometry. An optional single channel of EEG is included and can be used with a clear conductive gel without interfering with the NIR optics.

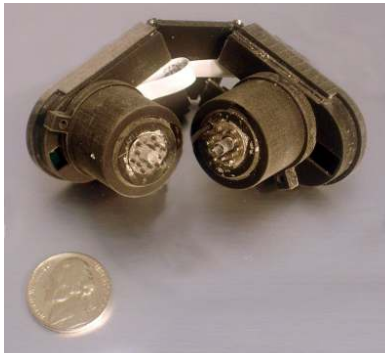


Fig. 1. Archinoetics' custom in-helmet near-infrared sensor

Design constraints focused on meeting flight requirements, including eliminating any potential tethering for communications or power, comfort for extended wear, and adhering to flight platform safety requirements, while maintaining signal integrity and strength.

Sensor electronics underwent miniaturization and power minimization in order to enable an in-helmet, wearable design. Custom optics used contributed to these modifications. A custom LED array was fabricated to emit the NIR light, requiring fewer mechanical components than lasers to stabilize the optical power and time-multiplex the three wavelengths used. Semi-custom silicon photodiodes with embedded gain and copper enclosure provides shielding to lower the noise level while still sensitive to even the minute changes associated with a hemodynamic response. Power minimization allowed for the system to be powered through a rechargeable, Li-Ion battery, eliminating the need for a connection to an external power source.

The inclusion of a 3-axis MEMS accelerometer, while adding little to the overall electronics footprint, provides valuable signals for environmental monitoring (such as the detection of high-G maneuvers during flight) as well as signals needed for further investigation into the detection and/or removal of motion artifacts that affect the NIR sensing signal.

The design of the sensor enclosure centered around subject comfort for extended wear and a form factor that would fit into a Navy flight helmet above the ear cup without compromising any of the safety features, while still maintaining the optical requirements for NIRS sensing. The resulting enclosure uses low-spring rate compression springs behind the optics to ensure constant but comfortable skin contact with the optical fiber penetrating hair for optics-skin coupling over any part of the head. Compressed spring height was balanced with the optical component and outer wall height for complete retraction of the optics to avoid skin penetration for flight impact safety. The optics enclosure is semi-spherically shaped with the rounded surface cradled against the top of the compression spring. This configuration allows self-rotation of the optical components to ensure an orthogonal coupling of the optics to the surface of the scalp as required of the optical path of the NIR light.

Fig. 2. Illustration with an internal view of the sensor enclosure design that fits within a Navy flight helmet above the ear cup with spring-mounted optics that allow for self-rotation of the optical components

Channels allow the optics components to move laterally and independently to ensure a 2.5 – 3.5 cm lateral separation between the emitter and detector for proper NIRS sensing of the cortical surface. The required spacing between the optics is dependent on the area of interest and individual differences in physiology. Torsion springs along the central joints apply low pressure against the imaging surface with the overall curvature of the sensor matching the curvature of the top of the ear cup when mounted within the helmet. Any electrical interconnects between components are done with sealed flex cables and travel through channels in the enclosure to protect the internal electronics against potential water or sweat contact.

To include the optional channel of EEG, an electrode is twisted into place on the top surface of the sensor enclosure. Any electrode can be used that is the same height as the optics, has a central void for the NIR optics and a conductive surface that extends to the back of the electrode. Electrical contact is made through a spring-loaded pin that is mounted on the internal circuit board and protrudes through the enclosure against the electrode base. This EEG channel has been used with Ag/AgCl coated probed and gold cup electrodes. Both types of electrodes were modified by reducing the height and adhering the electrode to a conductive washer with conductive epoxy.

Table 1. Features, advantages, and benefits of the custom NIR sensor created

Features	Advantages	Benefits
Combined EEG & NIR sensing	Multiple physiologic signal monitoring in a single package	Fewer devices to be worn
Wireless capabilities	No wiring required for data transmission	No wiring from helmet
Battery powered	External power source not required	No wiring between helmet and external power source. No tethering to cockpit power source.
Miniaturized	Small device form factor	Helmet integration without modifications to safety features

A USB module receives data wirelessly in the ISM 2.4GHz band from an external PC for the modularly-expandable C# software application. Additional layers of processing or display can be added onto the application using the XML-based scripting to include third-party modules in a processing pipeline. Included modules focus on tools for assurance of correct sensor placement, such as a raw light levels display and signal quality check through signal-to-noise ratios, graphical displays of the incoming signals, and data logging.

2 Case Study: Oxy-hemoglobin Response to Valsalva Maneuvers

The capabilities of the wearable sensor were demonstrated in a limited case study. Two subjects performed Valsalva maneuvers based on a protocol used in research published by Bluestone [8]. Raw NIRS optical signals were logged along with the 3-axis accelerometer data from the same sensor. The accelerometer data was used to verify the absence of significant movement which would cause movement artifacts in the optical signal. The NIRS signal was analyzed for the expected fluctuations in oxy-hemoglobin concentrations in response to the performance of Valsalva maneuvers.

2.1 Methods

Two subjects participated in this study, resulting in five trials total. Subjects were asked to minimize head movement during the protocol to alleviate the question of slippage artifact. Subjects rested (no task) for 60s, performed a Valsalva for a 30s task window, rested for 30s, performed a Valsalva for 30s, rested for 30s, performed a Valsalva for 30s, then rested for at least 90s. The first subject performed two trials as described and the second subject performed three trials for a total of five trials.

Subjects that participated are experienced divers that use Valsalva maneuvers while diving to equalize pressure in the ear and sinus cavities. They were instructed to perform the Valsalva maneuvers to the extent that they felt comfortable. Forceful Valsalvas for the entire 30s task window were performed by the first subject for his second trial.

The sensor was placed on the left temple of the subject and held in place with a Velcro strap. After sensor placement, signal quality was verified through the visual inspection of the raw optical signal for the presence of heartbeats. The sampling frequency used was 256Hz.

The raw optical near-infrared signal was logged for post-processing. These signals were converted to oxy-hemoglobin concentrations using a modified Beer-Lambert equation [9]. A median filter was applied to the raw optical and oxy-hemoglobin concentration signals to remove physiologic and high-frequency noise from the electronics. The median filter used a one-second window of past data in order to maintain the potential for real-time analysis in the future.

Raw accelerometer data was also logged at 1024Hz for detection of subject movement as a potential source of motion artifact in the optical signal.

2.2 Results

Subject raw optical data for the 760nm wavelength and oxy-hemoglobin data were converted into percent changes compared to a 30s baseline at the beginning of each

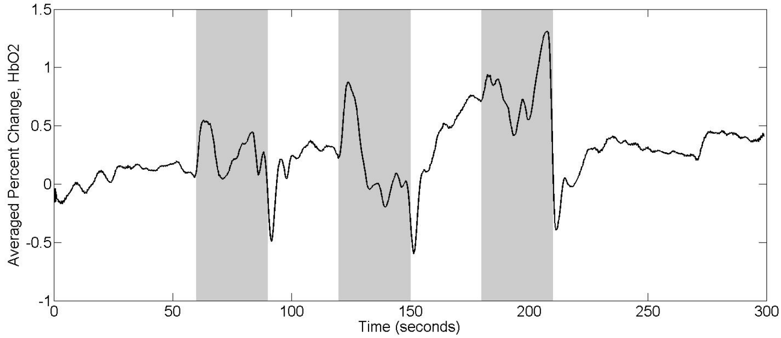


Fig. 3. Averaged oxy-hemoglobin concentration changes in response to Valsalva maneuvers (shaded periods)

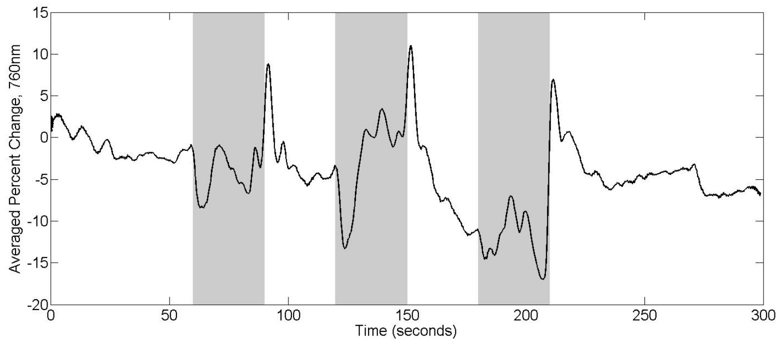


Fig. 4. Averaged 760nm wavelength optical intensity changes in response to Valsalva maneuvers (shaded periods)

trial. These signals were then averaged over the five trials, shown below. The periods when the subjects were performing the Valsalva maneuvers are indicated by the shaded areas.

Oxy-hemoglobin concentration increases ranged from to less than 1% to up to nearly 4% for the forceful Valsalvas performed by the first subject in his second trial. The 760nm signal shows the characteristic drop in optical intensity (essentially measured photons) during the maneuvers, but not to the same extent as in the Bluestone [8] study referenced due to the subjects choosing not to perform the Valsalva as forcefully as would normally be performed when not at 1g.

Data from the first subject's second trial, in which forceful Valsalvas were performed for the duration of each of the 30s periods, are shown in the figures below.

These data show a 4% increase in oxy-hemoglobin concentration and an approximately 40% decrease in optical intensity of the 760nm wavelength optical signal during the performance of the Valsalva maneuver.

The on-board accelerometer data revealed only small deviations from the 1*g* baseline with an rms error between the accelerometer output and the 1*g* baseline ranging

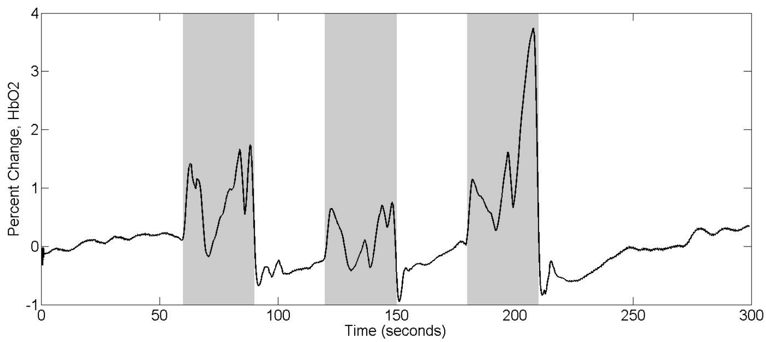


Fig. 5. Oxy-hemoglobin concentration changes in response to forceful Valsalva maneuvers (shaded periods), first subject, second trial

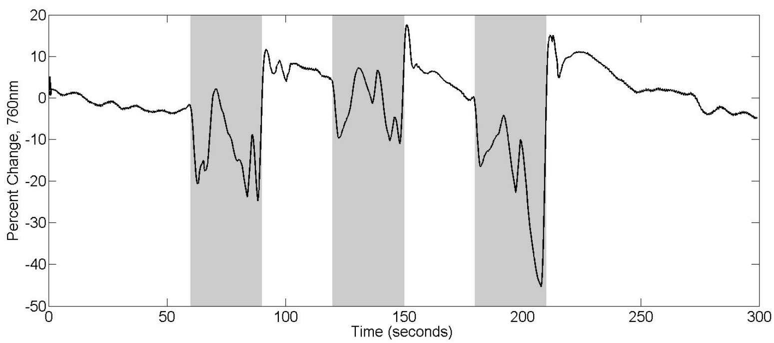


Fig. 6. 760nm wavelength optical intensity changes in response to forceful Valsalva maneuvers (shaded periods), first subject, second trial

from 0.072g to 0.083g (μ = 0.078g, σ = 0.005). Some of this deviation can be attributed to hardware noise, specified at 0.009g rms for the x- and y-axes and 0.011g rms for the z-axis for the 1024Hz sampling frequency used.

2.3 Discussion

Valsalva maneuvers involve the subject forcing expiration against a closed glottis which leads to an increase in arterial blood pressure and venous pressure, amongst other physiologic measures [10]. Several studies [11-14] have concluded that Valsalva maneuvers affect cerebral blood flow, and Bluestone et al suggests that these increases in cerebral blood flow are associated with an increase in cerebral oxy-hemoglobin concentrations [8].

The fluctuations observed in response to the Valsalva maneuvers performed by the subjects in this case study reflect the signal changes expected by the physiologic response to the maneuvers as shown in the Bluestone study replicated, although not to the same degree of signal change. The smaller signal change percentage in both the raw optical and oxy-hemoglobin concentration signals may reflect the subject-reported decisions not to perform the maneuvers as forcefully as could have been done. The more forceful maneuver performed by the first subject, with data shown in Figures 5 and 6, show signal changes of the same magnitude as the Bluestone study. Because the on-board accelerometer showed only small deviations from the 1g baseline, the signal changes observed should not be attributed to motion artifacts, but rather actual physiologic changes. The signal responses to these Valsalva maneuvers confirm both the near-infrared system functionality and its sensitivity to regional oxy-hemoglobin changes.

3 Conclusion

The custom NIR sensor features wireless operation in a wearable form factor that can be mounted within a flight helmet. Sensor outputs show expected responses to Valsalva maneuvers in the raw optical and oxy-hemoglobin concentrations signals.

Although the form factor of the sensor and its sensitivity to physiologic signals that may be early indicators of GLOC and hypoxia hint to flight applications, there are obstacles associated with integrating into a flight platform, including rigorous validation and strict regulatory approvals required. Other applications of the system designed are being researched, including physiologic measurements for clinical applications and cognitive activity research.

For these applications, helmet integration is not needed. An alternative form factor has been designed for making contact with the head without a helmet. Figure 7, below, shows a design for a functional brain imaging application, specifically for analysis of Broca's area where hair penetration for skin coupling is not required due to the location of that region. The enclosure holds a strap in place to aid in the coupling the sensor to the head without a helmet.

Fig. 7. Custom NIR sensor for verbal cognitive activity detection and analysis

Future work will focus on signal processing and algorithm development for artifact removal and extraction of signals of interest. These signals of interest include vital signs, such as heart rate, inter-beat intervals, and respiration rate derived from the oxy-hemoglobin concentration signal, as well as cognitive activations when applied to imaging of the cerebral cortex. Artifact removal is based on the application and signal of interest, with a large focus on motion artifact detection and/or removal (aided by the on-board accelerometer).

Additional efforts are being directed to the extraction of an accurate heartbeat signal from the NIR output. Physiologic signals unavoidably affect oxy- and deoxy-hemoglobin levels, with the dominant signals consisting of oscillations from the heartbeat and respiration. The heartbeat propagates changes in oxy- and deoxy-hemoglobin concentrations throughout the body tissue, resulting in changes in tissue transparency for NIR wavelengths outside of the isobestic point. Thus, the heartbeat signal is present in fNIR imaging as a pulse waveform, with the inflection points determined by the systolic pressure wave [15]. The use of the NIR signal for heartbeats would enable measurement from a single-site sensor, rather than multiple sensor sites as required by ECG-based measures.

References

1. Snow, M.P., Barker, R.A., O'Neill, K.R., Offer, B.W., Edwards, R.E.: Augmented cognition in a prototype uninhabited combat air vehicle operator console. In: Schmorrow, D.D., Stanney, K.M., Reeves, L.M. (eds.) Foundations of Augmented Cognition, 2nd edn., pp. 279–288. Strategic Analysis, Inc, Arlington (2006)
2. Ryoo, H.C., Sun, H.H., Shender, B.S., Hrebien, L.: Consciousness monitoring using near-infrared spectroscopy (NIRS) during high +Gz exposures. Medical Engineering and Physics 26(9), 745–753 (2004)
3. McKinley, R.A., Tripp Jr., L.D., Bolia, S.D., Roark, M.R.: Computer modeling of acceleration effects on cerebral oxygen saturation. Aviat. Space Environ. Med. 76(8), 733–738 (2005)
4. Lyons, T.J.: Analysis of mission and aircraft factors in G-induced loss of consciousness in the USAF: 1982-2002. Aviat. Space Environ. Med 75(6), 482–486 (2004)
5. Cable, G.G.: In-Flight Hypoxia Incidents in Military Aircraft: Causes and Implications for Training. Aviat. Space Environ. Med. 74(2), 169–172 (2003)
6. Nishimura, E.M., Stautzenberger, J., Robinson, W., Downs, T.H., Downs III, J.H.: A New Approach to fNIR: OTIS. IEEE Engineering in Medicine and Biology Magazine (July/August 2007)
7. Schmidt, F.E.W.: Development of a Time-Resolved Optical Tomography System for Neonatal Brain Imaging. Ph.D. Thesis, Dept. of Medical Physics and Bioengineering, University College London (1999)
8. Bluestone, A.Y., Abdoulaev, G., Schmitz, C.H., Barbour, R.L., Hielscher, A.H.: Three-dimensional optical tomography of hemodynamics in the human head. Optics Express 9(6), 272–286 (2001)
9. Franceschini, M.A., Boas, D.A., Zourabian, A., Diamond, S.G., Nadgir, S., Lin, D.W., Moore, J.B., Fantini, S.: Near-infrared spiroximetry: noninvasive measurements of venous saturation in piglets and human subjects. J. Appl. Physiol. 92, 372–384 (2002)
10. Tiecks, F.P., Lam, A.M., Matta, B.F., Strebel, S., Douville, C., Newell, D.W.: Effects of the Valsalva maneuver on cerebral circulation in healthy adults. Stroke 26, 1386–1392 (1995)
11. Matta, B.F., Lam, J., Smileswiski, P.: The effect of the Valsalva maneuver on cerebral hemodynamics: a near infrared spectroscopy study. J. Neurosurgical Anesthesiology 8, 601 (1996)

12. Pott, F., Van Lieshout, J.J., Ide, K., Madsen, P., Secher, N.H.: Middle cerebral artery blood velocity during Valsalva mameuver in the standing position. J. Appl. Physiol. 88, 1545–1550 (2000)
13. Dawson, S.L., Panerai, R.B., Potter, J.F.: Critical closing pressure explains cerebral hemodynamics during the Valsalva maneuver. J. Appl. Physiol. 86, 675–680 (1999)
14. Raisis, J.E., Kindt, G.W., McGillicuddy, J.E., Giannotta, L.: The effects of primary elevation of cerebral venous pressure on cerebral hemodynamics and intracranial pressure. J. Surg. Res. 26, 101–107 (1979)
15. Gratton, G., Corballis, P.M.: Removing the heart from the brain: Compensation for the pulse artifact in the photon migration signal. Psychophysiology 32, 292–299 (1995)

Assessment of Psychophysiological Differences of West Point Cadets and Civilian Controls Immersed within a Virtual Environment

Thomas D. Parsons[1,*], Christopher Courtney[1], Louise Cosand[1], Arvind Iyer[1], Albert A. Rizzo[1], and Kelvin Oie[2]

[1] University of Southern California's Institute for Creative Technologies. Marina del Rey, CA, USA
tparsons@ict.usc.edu
[2] U.S. Army Research Laboratory. Aberdeen Proving Ground, MD, USA

Abstract. An important question for ecologically valid virtual environments is whether cohort characteristics affect immersion. If a method for assessing a certain neurocognitive capacity (e.g. attentional processing) is adapted to a cohort other than the one that was used for the initial normative distribution, data obtained in the new cohort may not be reflective of the neurocognitive capacity in question. We assessed the psychophysiological impact of different levels of immersion upon persons from two cohorts: 1) civilian university students; and 2) West Point Cadets. Cadets were found to have diminished startle eyeblink amplitude compared with civilians, which may reflect that cadets experienced less negative affect during the scenario in general. Further, heart rate data revealed that Cadets had significantly lower heart rates than Civilians in the "low" but not "high" immersion condition. This suggests that "low" immersion conditions may not have the ecological validity necessary to evoke consistent affect across cohorts.

Keywords: virtual environment; psychophysiological assessment; immersion; ecological validity, neuropsychology.

1 Introduction

Neuropsychological studies tend to assess neurocognitive performance using standardized assessments in controlled settings and behavioral (i.e. self and other) rating scales for assessment of the subject's activities in a real-world setting. While traditional neuropsychological assessments manipulate the complexity of the stimulation, they do little to assess the impact of the intensity of the situation. Neuropsychological assessment should strive for ecologically valid assessments in which findings may reflect the varying levels of intensity found in real world situations. For example, simulations that proffer more intense presentations may elicit accompanying increase in emotional responses. In a related manner, findings from neuropsychological assessments should be generalizable to real-world situations [1]. While controlled

* Corresponding author.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 514–523, 2009.

settings offer increased psychometric rigor, naturalistic (i.e. observation-based) behavioral ratings may have increased ecological validity in that they may better capture the subject's performance in a real world setting. It is important to note that neuropsychological measures in controlled settings and behavioral ratings based upon naturalistic observations do not proffer consistently parallel findings. Further, dissimilar neurocognitive components may be dissociated both by neuropsychological measures in controlled settings and behavioral ratings based upon naturalistic observations [2].

1.1 Ecological Validity: Verisimilitude and Veridicality

Current approaches to increasing the ecological validity of neuropsychological assessments reflect the face validity of the assessments and the past experience of the neuropsychologist in drawing inferences between the results and the subject's activities of daily living. There are two approaches to the establishment of ecologically valid neuropsychological assessments: verisimilitude and veridicality [3]. From a human-computer interaction perspective, "verisimilitude" refers to the similarity between task demands of the test and demands imposed in the everyday environment. The establishment of verisimilitude requires "computer" tests that comprise the everyday cognitive tasks of the "human", such that inferences can be easily drawn from computer test results and the human's likely ability to perform those tasks in daily life. Contrariwise, "veridicality" of human-computer interaction refers to the extent to which results on a computer-based assessment are related to scores on other measures (e.g. standardized paper and pencil neuropsychological tests) that predict human performance of real-world tasks [3]. The establishment of veridicality requires that the researcher statistically assess the relationship between human performance on computer-based neuropsychological tests and measures of everyday functioning (e.g. behavioral observations and rating scales).

1.2 Virtual Environments for Neuropsychological Assessment

Virtual environments offer the capacity for merging the benefits of controlled settings (e.g. increased psychometric rigor) within ecologically valid virtual environments that simulate the naturalistic environment in which behaviors occur. Recent advances in simulation technology have produced new methods for the creation of virtual environments. With these systems, researchers can present users with ecological verisimilude reflective of "real world" environments. When delivered via a head-mounted display (HMD), an experience of immersion within these captured scenarios can be supported in human users. Immersion has been defined as one's subjective impression that she or he is participating in a comprehensive, realistic experience. Immersion may be enhanced by design strategies that increase the participant's suspension of disbelief [4]. As such, the VR assets that allow for precise stimulus delivery within simulations appear well matched for increasing the ecological validity of neuropsychological assessment.

The value in using virtual reality technology to produce simulations targeting neurocognitive and behavioral applications has been acknowledged by an encouraging body of research. Some of the work in this area has addressed affective processes: anxiety disorders, pain distraction, posttraumatic stress disorder [5]. Other work has assessed neurocognitive processes such as attention and executive functioning [6], [7]; memory [8], [9], [10]; and visuospatial abilities [11], [12], [13]. While multiple attempts have been made to apply theoretical perspectives to the development of believable virtual environments, little has been done to "objectively" assess human interpretations of immersion in these environments. There is need for the incorporation of psychophysiological metrics into assessment of human responses while immersed in a virtual environment. As mentioned above, neuropsychological assessment should aim to recreate the environment in which the subject will be processing information. This is especially important when persons are processing information while immersed in environments that have different levels of stimulus intensity. Exposure to emotionally intense situations results in regular activation of cerebral metabolism in brain areas associated with inhibition of maladaptive associative processes [14]. Identical neural circuits have been found to be involved in affective regulation across affective disorders [15], [16]. Systematic and controlled exposure to physiologically intense stimuli may enhance emotional regulation through adjustments of inhibitory processes on the amygdala by the medial prefrontal cortex during exposure and through structural changes in the hippocampus [17].

Thus far, the recording of psychophysiological variables while participants operate within virtual environments has produced useful results in studies examining immersion and presence [18], [19], [20]. As such, the VR assets that allow for precise stimulus delivery within ecologically enhanced scenarios appears well matched for this research. Researchers have found that the individual characteristics of study participants may impact the immersiveness and subsequent findings of a given study. Of primary importance is the extent to which a participant is capable of "absorption" and "hypnotism." Hence, individual differences may moderate presence and confound findings. The propensity of participants to get involved passively in some activity and their ability to concentrate and block out distraction are important factors to consider when conducting a study. Likewise, evidence suggests that hypnotizability plays a role in the outcome of studies using VR. Research into these moderating individual traits is of value because such research may augment participant selection [18].

An important question for ecologically valid virtual environments is whether cohort characteristics affect immersion. If a method for assessing a certain skill is adapted to other cohorts and therefore different for each cohort, data obtained in different cohorts are comparable only when changes to the environment designed for one cohort are not necessary to demonstrate the studied skill in the cohort concerned. For example, neuropsychological assessment of military personnel using a virtual environment that was norm referenced to civilians may not have the same predictive validity for both cohorts. Herein we assessed the psychophysiological impact of different levels of immersion upon persons from two cohorts: 1) civilian university students; and 2) West Point Cadets.

2 Methods

2.1 Participants

A total of 15 subjects participated in this experiment. Six subjects were West Point cadets and 9 subjects were civilian students and staff at the University of Southern

California. Strict exclusion criteria were enforced so as to minimize the possible confounding effects of additional factors known to adversely impact a person's ability to process information, including psychiatric (e.g., mental retardation, psychotic disorders, diagnosed learning disabilities, Attention-Deficit/Hyperactivity Disorder, and Bipolar Disorders, as well as substance-related disorders within two years of evaluation) and neurologic (e.g., seizure disorders, closed head injuries with loss of consciousness greater than 15 minutes, and neoplastic diseases) conditions. Subjects were comparable in age, education, ethnicity, sex, and self-reported symptoms of depression.

2.2 Procedure

The University of Southern California's Institutional Review Board approved the study. After informed consent was obtained, basic demographic information and computer experience and usage activities were recorded. While experiencing the VRCPAT, participant psychophysiological responses were recorded using the Biopac system. Following completion of the VRCPAT protocol, subjects completed the simulator sickness questionnaire, which includes a pre-VR exposure symptom checklist.

2.3 Virtual Reality Cognitive Performance Assessment Test

The project described herein builds upon a larger (ongoing) project that makes use of virtual environments to assess user sensory, perceptual, and neurocognitive performance on various tasks. Neurocognitive and psychophysiological data gleaned from such analyses provides opportunity for implementing systems that can exploit the capabilities of nervous systems, rather than simply depending upon human adaptation, to improve and optimize human-computer interaction. Monitoring the neurocognitive and psychophysiological activity of persons operating within a complex environment, however, poses severe measurement challenges. It is also likely that neurocognitive and psychophysiological responses in operational versus tightly controlled laboratory environments will be significantly, if not fundamentally, different than in controlled laboratory settings.

The VRCPAT project focuses on the refinement of neuropsychological assessment using virtual environments to assess persons immersed in ecologically valid virtual scenarios. The VRCPAT is a three-dimensional virtual environment (i.e. virtual city and Humvee scenarios) designed to run on a Pentium IV notebook computer with one gigabyte RAM and a 128 megabyte graphics card. The primary aim of the VRCPAT project is to use the already existing library of assets as the basis for creating a VE for the standardized assessment of neurocognitive performance within a contextually relevant VE. The application uses USC's FlatWorld Simulation Control Architecture (FSCA). The FSCA enables a network-centric system of client displays driven by a single controller application. The controller application broadcasts user-triggered or scripted-event data to the display client. The real-time three-dimensional scenes are presented using Numerical Design Limited's (NDL's) Gamebryo graphics engine. The content was edited and exported to the engine, using Alias's Maya software. Three-dimensional visual imagery is presented using the eMagin z800. Navigation through the scenario uses a common USB Logitech game pad device.

Virtual reality-based simulation technology approaches, as delineated herein, are considered to be the future alternative for devising neuropsychological assessment measures that will have better ecological/predictive validity for real-world performance. As well, the flexibility of stimulus delivery and response capture that are fundamental characteristics of such digital environments is viewed as a way for research objectives to be addressed in a more efficient fashion for long term needs. The overall design of this type of assessment tool allows for 1) Verisimilitude: the presentation of realistic environments that reflect activities of daily living; and 2) Veridicality: flexibility in terms of the independent variables that could be studied with this method once the psychometric properties of the standardized test are determined. Such flexibility enables this system to be viewed as an open platform on which a wide range of research questions may be addressed. These include the manipulation of: 1) information load on the front end via the intensity and complexity of target stimuli to be attended to and the type of information in terms of relevance, similarity, vagueness, sensory properties; 2) temporal constraints during varied sustained assessment conditions; 3) distracting activities during the neurocognitive assessments; 4) sensory modality of the information presentation that needs to be attended to; 5) the reward structure used during some tests to assess motivational factors that influence performance; 6) the presentation of aversive stimuli for stressed performance evaluations; and 7) the development of a test bed whereby neurocognitive training and augmented cognition strategies could be assessed under known conditions supported by normative standards.

2.4 Stimuli and Design

Subjects were immersed a virtual environment (VE) on two separate experimental runs consisting of both a “high” immersion condition and a “low” immersion condition. In the high immersion condition, subjects wore a head mounted display (HMD) with full tracking capabilities and were free to explore their environment visually. The high immersion condition also made use of headphones and a tactile transducer floor to simulate riding in a large vehicle. The low immersion condition consisted of the same virtual Iraqi scenario presented on a 17 inch laptop screen while wearing headphones. During the low immersion condition, subjects viewed the VE from a static position.

The VE was comprised of a series of safe and combat zones in an Iraqi city. In both the high immersion and low immersion conditions, subjects viewed the VE from the perspective of the driver of a Humvee. The speed and trajectory of the vehicle was kept constant to control for time spent in each zone of the VE. Safe zones consisted mainly of a road surrounded by a desert landscape and were free of gunfire and other loud noises. The combat zones included improvised explosive devices (IEDs), gunfire, insurgents, and screaming voices. Subjects passed through 3 safe and 3 combat zones on each experimental run. The total length of each run was 210 seconds.

An acoustic startle probe was used to elicit startle eyeblink responses. The startle probe was a 110 dB white noise burst 50 ms in duration with a near instantaneous rise/fall time presented binaurally through Telephonics TDH-50P headphones. Decibel levels were measured with a Realistic sound level meter using a Quest Electronics earphone coupler. Startle probes were experienced intermittently throughout the experimental runs. A total of 4 startle probes were experienced in both the safe and combat zones in each run.

2.5 Depcholendent Variables

Psychological Trait Assessment. The following measures were used to assess the impact of absorption and immersiveness upon the “believability” of the system. Prior to the experiment itself, the subjects were required to fill in the following questionnaires: 1) Tellegen Absorption Scale (TAS). The TAS questionnaire aims to measure the subject’s openness to absorbing and self-altering experiences. The TAS is a 34-item measure of absorption. 2) Immersive tendencies questionnaire (ITQ). The ITQ measure individual differences in the tendencies of persons to experience “presence” in an immersive VE. The majority of the items relate to a person’s involvement in common activities. While some items measure immersive tendencies directly, others assess respondents’ current fitness or alertness, and others emphasize the user’s ability to focus or redirect his or her attention. The ITQ is comprised of 18 items, and each is rated on a 7-point scale.

Psychophysiological Assessment. Psychophysiological assessment included: Electromyographic activity (EMG), Electrodermal activity (EDA), Electrocardiographic activity (ECG), and respiration, which were recorded simultaneously using a Biopac MP150 system and a computer running Acknowledge software.

Startle eyeblink response. EMG startle eyeblink responses were recorded using two small (4mm in diameter) silver-silver chloride electrodes placed over the orbicularis oculi muscle of the left eye and an 8mm silver-silver chloride electrode placed behind the left ear to serve as a ground. One 4mm electrode was placed directly below the pupil in forward gaze while the other was placed about 1 cm lateral to the first. The electrodes were placed as close to the eye as possible while still allowing the subject to open and close his or her eyes comfortably. Impedance between the two electrodes was measured and deemed acceptable if below 10 kΩ.

The raw EMG signal was recorded at a rate of 1000 Hz throughout the experiment using a 10 Hz high pass and 200 Hz low pass filter. Raw signals were stored and exported for analysis in microvolt (μV) values.

The raw EMG signal was rectified and integrated for analysis. In order to qualify for scoring, the eyeblink trace had to begin within a window of 20 to 100 ms following the offset of the startle probe. The eyeblink response had to reach peak activity within a window of 20 to 150 ms following the startle probe. Blinks occurring at longer latencies were not considered to be the result of the startle probe. Amplitudes were recorded as the difference between the peak activity value and the baseline level present immediately preceding onset of the blink response. If the subject was blinking during the onset of the startle probe, that blink response was removed from further analysis due to artifact.

Cardiovascular responding. ECG was recorded with use of a Lead 1 electrode placement, with one 8 mm silver-silver chloride electrode placed on the right inner forearm about 2 cm below the elbow and another placed in the same position on the left inner forearm. A third 8 mm silver-silver chloride electrode was placed on the left inner wrist to serve as a ground. Electrode sites were cleaned with alcohol prep pads in order to improve contact. Interbeat intervals (IBIs) were scored as the time difference in seconds between successive R waves in the ECG signal. A median interbeat interval was recorded during each of the same 5 second sampling periods used to assess skin conductance level.

3 Results

3.1 Data Analytic Considerations

First, we assessed the potential impact of psychological characteristics such as absorption and immersiveness upon the "believability" of the virtual environment. No significant differences were found between groups on these measures. After controlling for potential confounds related to absorption and immersiveness, physiological data were processed using custom-written programs. Within each subject, the median data point for each measure and condition was selected for analyses. Assumptions of a normal distribution equate the mean and median, and the median is less sensitive to outliers. As physiological data are susceptible to introduction of artifact from sources both inside of and outside of the body, the median is equivalent to the mean under ideal circumstances and superior to the mean when artifact alters some data points. Mean-based comparisons between subjects' medians, however, were considered appropriate because artifact-laden data points were already filtered out. A series of ANOVAs were performed on psychophysiological results to assess impact of cohort (Cadet vs Civilian) and level of immersion (High Immersion vs Low Immersion). First analysis of variance was performed on startle eyeblink amplitudes and next upon participants' heart rates, as measured in interbeat intervals.

3.2 Startle Eyeblink Amplitudes

A series of ANOVAs were performed. First, a two (Cadet vs Civilian) by two (High Immersion vs Low Immersion) ANOVA was performed on startle eyeblink amplitudes. This analysis revealed a main effect of group membership as Cadets' startle eyeblink amplitudes were smaller than those of Civilians ($F= 7.249$, $p< 0.05$). While Cadets had lower startle amplitudes overall, the difference was more exaggerated in the High Immersion condition ($F= 4.695$, $p < 0.05$). As increased startle eyeblink amplitude is associated with negative affect, these data suggest that Cadets were less emotionally impacted by the experience than were Civilians.

3.3 Heart Rates Measured in Interbeat Intervals

Participants' heart rates, as measured in interbeat intervals, were also analyzed using a two (Cadet vs Civilian) by two (High Immersion vs Low Immersion) ANOVA. When comparing interbeat intervals between groups a strong main effect occurs in the Low Immersion condition. In the Low Immersion condition, Cadets show significantly larger interbeat intervals, meaning a lower heart rate, than Civilians ($F= 17.662$, $p < 0.05$). The effect was not seen for the High Immersion condition ($F= 0.001$, $p= 0.997$), suggesting that both Cadets and Civilians were responding to the scenario in a similar way.

4 Discussion

Neuropsychological assessment should strive for ecologically valid assessments in which findings reflect the varying levels of intensity found in real world situations. While controlled settings offer increased psychometric rigor, naturalistic behavioral ratings may have increased ecological validity. Unfortunately, neuropsychological measures in controlled settings and behavioral ratings based upon naturalistic observations do not proffer consistently parallel findings. Although virtual environments offer the capacity for merging the benefits of controlled settings (e.g. increased psychometric rigor) within ecologically valid virtual environments that simulate the naturalistic environment in which behaviors occur, the question of whether cohort characteristics affect immersion remains to be established. Herein we assessed the psychophysiological impact of different levels of immersion upon persons from two cohorts: 1) civilian university students; and 2) West Point Cadets.

West Point Cadets were found to have diminished startle eyeblink amplitude compared with civilian controls. As eyeblink amplitude is thought to increase as negative affect increases, so the cadets' relatively small startle eyeblink responses suggest that cadets experienced less negative affect during the scenario in general. For example, there were not significant differences in the Cadets' eyeblinks for level of immersion (i.e. high versus low) or level of scenario intensity (i.e. safe versus ambush zones). Hence, it was difficult to differentiate the impact of immersion upon both cohorts. For both the low and high immersion conditions, Civilians' eyeblink amplitudes were significantly greater than that of Cadets, suggesting that Cadets were well accustomed to experiences with high levels of intensity.

The possibility exists, however, that these data may be interpreted in terms of attention. The startle reflex is often modulated by attentional processing, with more focused attention correlating with a decrease in startle eyeblink amplitude. Findings within the Cadet cohort may reflect their more intense focus on the task, which may have lead to decreased startle responses. Further, it is important to note that startle eyeblink amplitudes in this study were most likely modulated by both negative affect and attention. Unfortunately, the current design does not disentangle the unique contributions of each. Future studies should make use of an attentional processing paradigm to support the differentiation of these components. Cardiac data, however, corroborates the interpretation that negative affect is an important source of variance between Cadets and Civilians.

Heart rate data, collected using interbeat intervals, found that Cadets had significantly lower heart rates than Civilians in the low immersion condition. In the high Immersion condition, however, Cadets' cardiac responses resembled those of Civilians. This suggests that low immersion was insufficient to evoke negative affect in Cadets, but that high immersion impacted the Cadets' affective system. These data also serve to support the affective, versus attentional, interpretation of the startle eyeblink data. If attention were modulating responses there would be equivalent heart rate between the two levels of immersion or a drop in heart rate during high immersion because high immersion engages more attention. The psychophysiological findings of this study suggest that level of immersion is important for Cadet training in order to impact the Cadets' affective system.

In summary, ecologically valid virtual environments may require increased attention to cohort characteristics. If a method for assessing a certain neurocognitive capacity (e.g. attentional processing) is adapted to a cohort other than the one that was used for the initial normative distribution, data obtained in the new cohort may not be reflective of the neurocognitive capacity in question. The psychophysiological impact of different levels of immersion upon persons from military and civilian cohorts revealed cohort differences suggesting that "low" immersion conditions may not have the ecological validity necessary to evoke consistent affect across cohorts. It is important to note that psychophysiological results in this study were most likely modulated by attentional processing. Unfortunately, the current design does not disentangle the unique contributions of attentional processing. Future studies should make use of an attentional processing paradigm in a larger sample to support the differentiation of these components.

References

1. Foa, E.B., Kozak, M.J.: Emotional processing of fear: exposure to corrective information. Psychological Bulletin 99, 20–35 (1986)
2. Gordon, M., Barkley, R.A., Lovett, B.J.: Tests and observational measures. In: Barkley, R.A. (ed.) Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment, 3rd edn., Guilford, New York, pp. 369–388 (2006)
3. Spooner, D.M., Pachana, N.A.: Ecological validity in neuropsychological assessment: A case for greater consideration in research with neurologically intact populations. Archives of Clinical Neuropsychology 21, 327–337 (2006)
4. Dede, C.: Immersive Interfaces for Engagement and Learning. Science 323, 66–69 (2009)
5. Parsons, T.D., Rizzo, A.A.: Affective Outcomes of Virtual Reality Exposure Therapy for Anxiety and Specific Phobias: A Meta-Analysis. Journal of Behavior Therapy and Experimental Psychiatry 39, 250–261 (2008)
6. Parsons, T.D., Bowerly, T., Buckwalter, J.G., Rizzo, A.A.: A controlled clinical comparison of attention performance in children with ADHD in a virtual reality classroom compared to standard neuropsychological methods. Child Neuropsychology 13, 363–381 (2007)
7. Parsons, T.D., Rizzo, A.A.: Neuropsychological Assessment of Attentional Processing using Virtual Reality. Annual Review of CyberTherapy and Telemedicine 6, 23–28 (2008)
8. Parsons, T.D., Rizzo, A.A., Bamattre, J., Brennan, J.: Virtual Reality Cognitive Performance Assessment Test. Annual Review of CyberTherapy and Telemedicine 5, 163–171 (2007)
9. Parsons, T.D., Rizzo, A.A.: Initial Validation of a Virtual Environment for Assessment of Memory Functioning: Virtual Reality Cognitive Performance Assessment Test. Cyberpsychology and Behavior 11, 17–25 (2008)
10. Parsons, T.D., Silva, T.M., Pair, J., Rizzo, A.A.: A Virtual Environment for Assessment of Neurocognitive Functioning: Virtual Reality Cognitive Performance Assessment Test. Studies in Health Technology and Informatics 132, 351–356 (2008)
11. Parsons, T.D., Larson, P., Kratz, K., Thiebaux, M., Bluestein, B., Buckwalter, J.G., Rizzo, A.A.: Sex differences in mental rotation and spatial rotation in a virtual environment. Neuropsychologia 42, 555–562 (2004)

12. Parsons, T.D., Rizzo, A.A., Buckwalter, J.G.: Backpropagation and regression: comparative utility for neuropsychologists. Journal of Clinical and Experimental Neuropsychology 26, 95–104 (2004)
13. Parsons, T.D., Rizzo, A.A., van der Zaag, C., McGee, J.S., Buckwalter, J.G.: Gender and cognitive performance: a test of the common cause hypothesis. Aging, Neuropsychology, and Cognition 12, 78–88 (2005)
14. Schwartz, J.M.: Neuroanatomical aspects of cognitive-behavioural therapy response in obsessivecompulsive disorder. An evolving perspective on brain and behaviour. British Journal of Psychiatry Supplemental, 38–44 (1998)
15. De Raedt, R.: Does neuroscience hold promise for the further development of behavior therapy? The case of emotional change after exposure in anxiety and depression. Scandinavian Journal of Psychology 47, 225–236 (2006)
16. Mineka, S., Watson, D., Clark, L.A.: Comorbidity of anxiety and unipolar mood disorders. Annual Review of Psychology 49, 377–412 (1998)
17. Hariri, A.R., Bookheimer, S.Y., Mazziotta, J.C.: Modulating emotional responses: effects of a neocortical network on the limbic system. Neuroreport 11, 43–48 (2000)
18. Macedonio, M., Parsons, T.D., Rizzo, A.A.: Immersiveness and Physiological Arousal within Panoramic Video-based Virtual Reality. Cyberpsychology and Behavior 10, 508–516 (2007)
19. Meehan, M., Insko, B., Whitton, M., Brooks, F.: Physiological measures of presence in virtual environments. In: Proceedings of 4th Annual Presence Workshop, Philadelphia (May 2002)
20. Pugnetti, L., Meehan, M., Mendozzi, L.: Psychophysiological correlates of virtual reality: a review. Presence 10, 384–400 (2001)

Characterizing the Psychophysiological Profile of Expert and Novice Marksmen

Nicholas Pojman[1], Adrienne Behneman[1], Natalie Kintz[1], Robin Johnson[1], Greg Chung[2], Sam Nagashima[2], Paul Espinosa[2], and Chris Berka[1]

[1] Advanced Brain Monitoring, Inc. 2237 Faraday Ave Ste. 100, Carlsbad, CA 92008
chris@b-alert.com
[2] UCLA/National Center for Research on Evaluation, Standards, and Student Testing. 10945 Le Conte Ste 1400 Mailbox 957150 Los Angeles, CA 90095-7150
greg@ucla.edu

Abstract. Marksmanship training includes a combination of classroom instruction and field practice involving the instantiation of a well-defined set of sensory, motor, and cognitive skills. 10 expert marksmen and 30 novices participated in a study that measured marksman performance during simulated ballistics shooting of a M4 replica infrared rifle. Participants' physiology and performance were quantified while they completed a battery of neurocognitive tests. Experts demonstrated consistent and more accurate shot performance across all trials. Compared to novices, experts evidenced lower levels of sympathetic activation as measured by heart rate variability during the neurocognitive tasks. Factor analysis identified experts as having above normal visuospatial processing speeds and sustained attention, reflecting experts as having better performance during vigilance neurocognitive tasks. Identifying physiological metrics of experts during neurocognitive testing opens the door to individualized novice instruction to help to improve specific areas flagged as below normal during or prior to novice marksmanship instruction.

Keywords: Electroencephalogram (EEG), Electrocardiogram (EKG), Marksmanship, Expert, Heart Rate Variability, Neurocognitive testing, psychomotor skill acquisition.

1 Introduction

Rifle marksmanship is a core skill for multiple branches of the armed forces; many members are required to qualify annually. Marksmanship training is generally a two-week program and includes a combination of classroom instruction and field practice involving instantiation of a well-defined set of sensory, motor, and cognitive skills. This translates into an estimated 352,000 person-weeks per year making individualized coach supervision and instruction difficult. Instruction and development of the fundamentals of marksmanship demands a large amount of resources from the military. Integrated neuroscience-based evaluation technologies coupled with targeted pre-training interventions could provide quantitative markers of successful learning

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 524–532, 2009.

and accelerate marksmanship skill acquisition, thereby saving military resources and potentially improving the safety and efficacy of troops. Furthermore, the rapid acquisition of expertise in marksmanship could serve as a model for aiding the instruction of other physical or mental skills required in military or other educational environments**.**

Marksmanship is a complicated psychomotor skill that demands high physical and mental coordination for proper execution, and can thus be conceptualized as a complex skill. Proper rifle shooting is accomplished through the synchrony of breathing; gross motor control of body positioning; fine motor control of muzzle wobble and trigger pull; and the processing of rear and front sight alignment with respect to a target. Psychomotor skill acquisition is commonly viewed as a progression from the initial cognitive phase in which new knowledge is assembled; intermediate associative phase where newly learned skills are slowly automated with practice; and a final autonomous phase in which task execution is automated and demands minimal conscious effort [1, 2]. Chung et al. (2006) used this phasic model of skill acquisition as a framework to identify five variables of rifle marksmanship that influence the development from early to final phases: perceptual-motor, cognitive, affective, equipment, and environment. Marksmanship expertise is accomplished through controlling these variables through the practice, development, and automation of fundamental marksmanship skills.

Previous EEG studies have revealed notable neurophysiological distinctions between expert and novices marksman during shooting [3-5]. EEG patterning of skilled marksmen during the preparatory shooting (pre-shot) period has identified distinct alpha and beta wave patterns between executed (fired) and rejected (non-fired, withdrawn) shots [6]. Several other studies have identified unique cortical activation patterns associated with successful shots in experts [7] and demonstrated distinct patterns between experts and novices [8]. Prior research comparing EEG during shooting to EEG during comparative novel visuospatial and visuo-analytic tasks found no significant differences between novices and experts during the two visual tasks to which neither group had been exposed before. However, researchers did discover unique alpha, beta, and gamma power differences during shooting. This unique EEG pattern has been theorized to be an indicator of experts' increased efficiency during only those visuospatial tasks in which they held expertise [9].

Heart rate is an additional physiological measure commonly analyzed in marksmanship studies. The preparatory period, or pre-shot period, is characterized by a heart rate deceleration and decrease in electrodermal skin conductance levels [10, 11]. Heart rate deceleration during the pre-shot period in experts may reflect the attentive and the skill-related aspects of sensory-motor preparation for performance [12]. These data indicate that both EEG and HR are measures that previous research has used to link physiological patterns to performance in experts marksmen. In addition to these factors, assessing physiological metrics during non shooting neurocognitive testing may provide reliable predictors of marksmanship expertise. Identification of cognitive and physiological metrics that distinguish expert from novice marksmen may prove useful in early identification of a novice psychophysiological state which may hinder or prevent progression towards expertise. This identification offers the possibility of early triage and potential intervention that may be tailored to meet the unique needs of each individual.

2 Methods

10 qualified expert marksmen (off-duty military) and 30 novices were recruited and screened for general health. Experts were all male, with an age range of 21-27 years (M= 23.7, SD=1.78). Novices were comprised of UCLA students and volunteers from the general population with no prior marksmanship experience. Novice participants (9 female, 21 male) ranged in age from 18 to 41 years (M=23.17, SD=4.80) and were screened to verify minimal shooting experience.

Ballistics simulation relied on a LaserShot rifle simulation set-up. A LaserShot M4 rifle trainer with simulated recoil replicates the weight (~ 8 lbs), sound, and kick of a real weapon using a CO_2 pneumatic recoil system. The weapon is fitted with an infrared (invisible) laser that "fires" upon trigger pull. Shooting performance was captured using the rifle's built-in infrared laser, infrared camera, and a digital projector. Each laser "shot" was detected by an Ultra Series Laser detection calibrated using LaserShot Ultra Series Camera Software. The Ultra Series Camera generates a mouse-click using the location of an infrared laser strike on the projected screen. These clicks were interpreted by computer to yield x and y coordinates on the target screen. A data integration program, Fusion 4000, was developed by collaborators at UCLA-CRESST to allow for a streamlined data collection. This software scaled targets and captured shooting performance. Data was collected in rooms at least 8.3 yards (~300") long. The digitally projected target was scaled by measuring the width of the projected screen and the screen's distance to shooter using. Fusion 4000 then adjusted the displayed target to simulate a 20" target at 200 yards [13].

All participants were asked to shoot from the kneeling position, one of the four required qualifying shooting positions [14]. Expert participants were given no instruction or feedback and completed 5 trials of 5 shots. Novices completed at least 8 trials of 5 shots each. After each trial, novices were shown the locations of their shots on a computer monitor along with number indicators next to each shot so they could identify when each strike occurred. All shots were fired at the participants will with no time restraint imposed on either the Experts or Novices. The pacing of shots was not regulated for any trial. All novices received basic instruction on proper positioning, handling, and sight alignment of the weapon prior to trial 1. Initial instruction was given by the same researcher for all novices. Novice participants were asked to shoot from the kneeling position but were allowed to choose any of the three variations of the kneeling position: high, medium, and low kneeling positions. The researcher modeled each of these positions and used posters of right and left handed versions of each position for additional reference. After the first two trials, novices watched a 15 minute video of a qualified marksmanship coach providing further details regarding proper shooting and positioning techniques.

A second session for collection of the neurocognitive test battery was completed on a separate visit, either before or after the shooting visit. All participants were given a battery of neurocognitive tests using the patent-pending Attention and Memory Profiler (AMP^{TM}) system. The AMP test battery included 3-Choice-Vigilance-Test (3C-VT), Standard Image-Recognition (SIR), Image Recognition with Interference (IIR), Verbal Paired-Associate-Learning (VPA), Number Image Recognition (NIR), Sternberg-Verbal-Memory-Scan (VMS), Eyes Open timed vigilance task (EO), and Eyes Closed (EC) timed vigilance task. These tasks were completed by participants using a computer in an isolated room. Participants' EEG, EKG, and task performance were measured while they completed visual memory and vigilance tasks.

EEG was recorded during both the marksmanship and neurocognitive sessions using the wireless B-Alert® 6-channel differential EEG headset [Figure 1]. The headset was designed with fixed sensor locations for three sizes (e.g., small, medium and large) with placement determined according to the International 10 – 20 system coordinates. All participants wore a 6 channel headset with 6 electrode sites at F3, C3, C4, Fz, Cz, and POz. Heart rate (HR) was measured with two sensors placed over the left collarbone and under the fifth lower right rib. Four neurocognitive factors are derived from the AMP data: sustained attention, processing speed, verbal memory and visuospatial memory, each with a quantitative, normalized score (based on our database of over 300 healthy data sets). Alertness, attention, verbal/visuospatial learning, and memory are also quantified using a combination of EEG and performance metrics. HR is recorded and quantified using PSD [14] to explore the relationship of sympathetic activation during AMP tasks with marksmanship performance.

Heart rate variability (HRV) was computed by first detecting QRS complexes in the EKG, and using the distance between them to calculate heart rate (raw HR). The raw HR is interpolated so that instead of a grid defined by heart beats, the grid is defined by seconds (HR). Each 5 minute segment of HR signal is then modeled as a 25^{th} order AR process, the coefficients of the process are estimated from the data and used to calculate the power spectra in the range from 0.001 - 0.05Hz in steps of 0.001Hz. Low frequency (LF) HRV is equal to the sum of power spectrum from 0.04 – 0.15Hz. High Frequency (HF) HRV is the calculated as the sum of power spectrum from 0.15 – 0.4Hz. These data were then z-scored to our healthy, fully rested database subjects.

The marksmanship performance measure was shot group *precision*. We defined shot group precision as the mean distance of each shot from the center of the shot group, where lower values reflect better precision. Use of shot group precision is a useful measure to assess shooter consistency as well as accuracy. In general, experts' shots were closer to the center of the shot group compared to novices, suggesting tighter shot groups.

Neurocognitive factor scores (NCFS) are comprised of four composite variables (factors): Visuo-Spatial Processing Speed (VSPS), Sustained Attention (SA), Recognition Memory Accuracy (RMA) and Recognition Processing Speed (RPS). The

Fig. 1. 6 Channel B-Alert® (*right*) system wirelessly records EEG and EKG. AMP set-up (*left*) measures performance and physiology during a battery of neurocognitive tests.

factors are derived from the measures of behavioral performance (i.e. mean reaction time, and percent of correct responses) during the various AMP™ tasks: Verbal Memory Scanning test, standard PAL test, verbal PAL test, interference PAL, numbers PAL, whole 3CVT and quartiles (0-5, 5-10, 10-15 and 15-20min) of the 3CVT (a total of 20 primary measures). Each raw factor score is then z-transformed with respect to the mean and standard deviation of the same score in a large reference population of normal subjects. Negative values indicate poor performance whereas positive values indicate good performance, with scores less than -2 indicated severe deficits.

3 Results

Shooting performance was determined using shot group precision on a trial by trial basis. Novice trials were summarized at four time points to reflect improvement as novices had more practice and instruction [Figure 2]. The worst novice performance was seen at baseline (n=30 participants X 2 trials, M=16.63, SD= 14.97), which was calculated as the mean distance to shot group center from trials 1&2. Trial 3 (M=15.77, SD= 11.93) included only the trial immediately following video instruction (Trial 3). Final (M=12.59, SD=8.54) included trials 7 and 8 for novices and showed the greatest shot group precision, thus reflecting the best average performance for novices. Expert trials were grouped into two groups, baseline (trials 1 &2) and final (trials 4&5). Expert marksmen showed little change from baseline (M=4.23, SD=2.70) to final (M=4.28, SD=2.54) trials. The standard deviation (variance) of experts' shots is also smaller, reflecting a higher consistency of shots for experts than novices.

Mann-Whitney U-test was used to compare novices to experts at baseline and final, respectively, do to uneven n and unequal variance of the samples. This analysis found that the Novices were significantly worse both at baseline (U=9, Experts=9, Novices=30, $p < 0.0001$) and final (U=5, Experts=9, Novices=28, $p < 0.0001$) time points, see Figure 2. Novice's improvement over time was examined with RMANOVA comparing baseline trials to final. RMANOVA found a significant improvement over time for the novice subjects [F (1,27)= 5.953, $p < .05$). These data are shown in Figure 3. Expert performance did not change over time ($p > .05$).

Neurocognitive factor analysis revealed significant performance-based changes between expert and novice groups in VSPS and SA factors [Figure 3]. Novices were randomly down-selected from n=28 to n=15 for ANOVA comparison of VSPS, SA, RMA, and RMS between novice and expert participants in order to meet the parameters required for ANOVA. A t-test was used to compare the subject included in these analysis to those that were excluded and found no significant difference (p=.487). Expert marksmen had a mean VSPS (M=2.12,) nearly 2 SD greater than the AMP normative database. Our novice shooters also had higher VSPS (M=0.65,) than the normative database. While both groups had VSPS above normal, experts did show

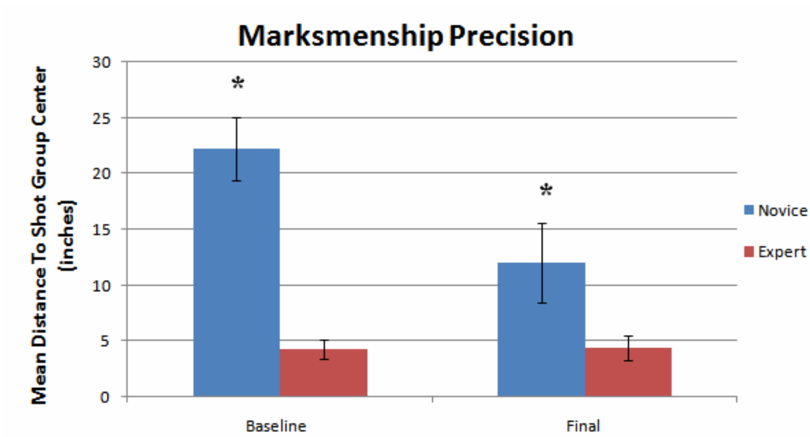


Fig. 2. Experts performed significantly better at both *Baseline* (Trials 1&2) and *Final* (Trials 7&8) compared to novices. Asterisk (*) indicates p<.01.

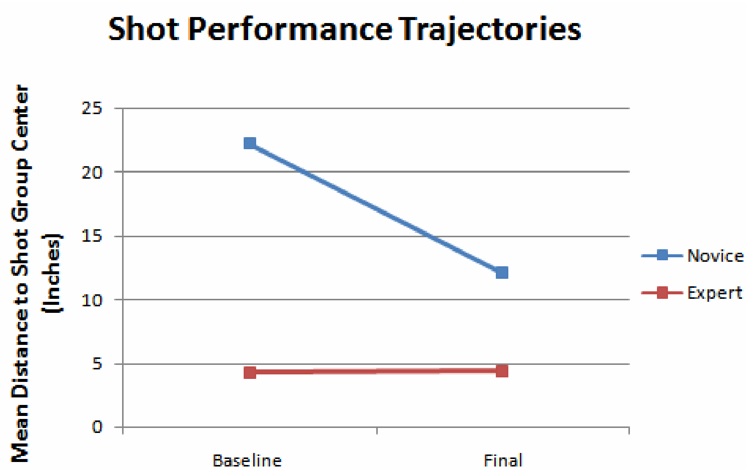


Fig. 3. Novice performance generally improved as they progressed through the trials. Mean distance to shot group center significantly improved from *Baseline* to *Final* Experts showed no change in performance from *Baseline* (Trials 1&2), to *Final* (Trials 4&5).

significantly higher VSPS as compared to novices according to ANOVA [F (1,23) = 6.13, p < .05]. ANOVA also found that experts had significantly higher SA [F (1,23) = 9.686, p <. 01] compared to novices although more than 1.5*SD* below the normative database.

Both experts and novices heart rate variability means were z-scored to a database of normal, healthy subjects. ANOVA analysis showed lower sympathetic associated (LF) HRV in the low demand tasks EO and EC for the experts compared to the novices; EO [F (1,23) = 5.061, p <.05], and EC [F (1,23) = 4.817, p <.05]. No group differences in HRV were found for any other tasks.

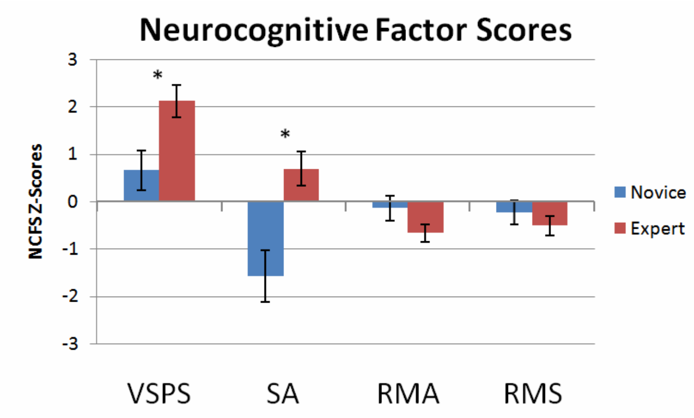


Fig. 4. Experts showed significantly higher scores for both Visuospatial Processing Speed (VSPS) and Sustained Attention (SA). No significant differences were seen in recognition memory accuracy (RMA) and recognition memory speed (RMS). Asterisk (*) indicates p<.05.

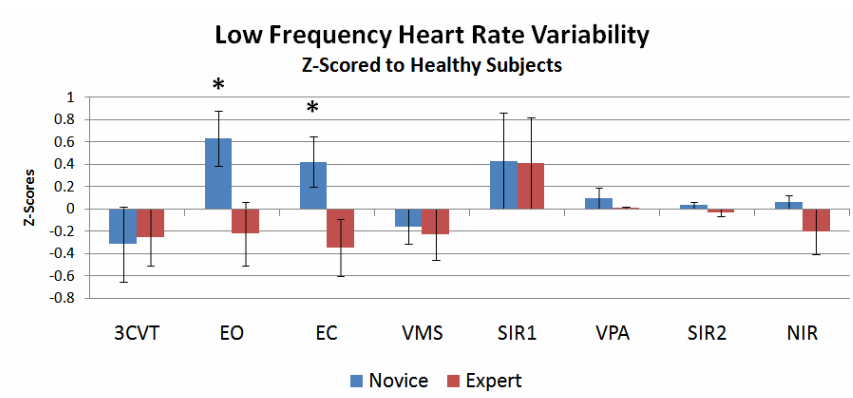


Fig. 5. Novices had significantly higher LF HRV compared to experts, reflecting higher anxiety in novices during timed vigilance tasks (3C-VT, EO, and EC). Asterisk (*) indicates p<.05.

4 Discussion

Assessment of shot group accuracy as measured by mean distance to shot group center shows that experts performed significantly better than novices across all trials. Experts showed no significant change from baseline to final shot trials reflecting the expert's controlled and consistent ability to replicate exceptional performance within and between trials. Consistent shot performance as indicated by no change across expert trials demonstrates that experts have the true ability to consistently apply the fundamentals of marksmanship. Neurocognitive factor analysis revealed experts as having VSPS more than 2 *SD* greater than normal levels and SA substantially above normal reflecting better performance on vigilance tasks. SA and VSPS neurocognitive factors are heavily derived from performance on the vigilance tasks (3C-VT, EO, and EC).

Experts had lower z-scored LF HRV for EO and EC tasks, compared to Novices. Greater LF HRV for novices may reflect increased sympathetic activation possibly indicating increased anxiety and mental stress during these neurocognitive vigilance tasks [15, 16]. It is important to note that HRV distinctions between experts and novices were only significantly different during the least cognitively challenging tasks in the AMP neurocognitive battery, the two timed vigilance tasks (EO and EC). This may indicate that experts are more able to regulate their cardio-respiratory function in a task specific/task appropriate manner. Further investigation comparing HRV during vigilance tasks and HRV during shooting may find distinct contrasts in HRV change between expert and novice group.

It is unclear whether this apparent ability to regulate expert's physiology is a genetically determined trait or a skill that can be acquired and refined with training. Our results highlight the need to investigate improving visuospatial processing speed, attention and cardiovascular regulation as a way to potentially improve novice marksmanship performance by early intervention designed to move novices toward the psychophysiological state observed in expert marksmen. Novices with below average VSPS may benefit from training using first person video gaming environments, which research has shown to improve visuospatial speed and attention [17]. Heart rate biofeedback or relaxation training may help novices develop control of sympathetic and parasympathetic activation to reduce anxiety and stress during vigilance tasks [15, 16, 18].

These results confirm research that asserts experts are capable of modulating their physiology to appropriately match task demands. Targeting physiological areas of weakness may increase novice's ability to automate the psychophysiological skills needed during shooting. Our results suggest the need for further research investigating whether pre-training interventions aid in the acceleration of learning for novices. These metrics may also have applications in other areas of psychomotor skill acquisition including additional military and education environments.

Acknowledgments. This work was supported by NIH\NIMH Grant No. MH078436, NIH\NHLBI Grant No. HL70484, and DARPA\DSO Grant No. 20064169. Author Berka is a shareholder in Advance Brain Monitoring, Inc.

"A" (Approved for Public Release, Distribution Unlimited).

References

1. Chung, G.K.W.K., et al.: New Directions in Rifle Marksmanship Research. Military Psychology 18(2), 161–179 (2006)
2. Fitts, P.M., Posner, M.I.: Human Performance. Brooks/Cole, Belmont (1967)
3. Hatfield, B., Hillman, C.: The psychophysiology of sport: A mechanistic understanding of the psychology of superior performance. In: Singer, R., Hausenblas, H., Janelle, C. (eds.) Handbook of Sports Psychology, pp. 362–386. Wiley & Sons, New York (2004)
4. Hatfield, B., et al.: Electroencephalographic studies of skilled psychomotor performance. Journal of Clinical Neurophysiology 21, 144–156 (2004)

5. Milton, J., et al.: Expert motor performance: Limbic activation is inversely related to the skill level. Neurology 60(Suppl. 1), A345 (2003)
6. Hillman, C., et al.: An electrocortical comparison of executed and rejected shots in skilled marksmen. Biological Psychology 52(1), 71–83 (2000)
7. Loze, G., Collins, D., Holmes, P.: Pre-shot EEG alpha-power reactivity during expert air-pistol shooting: a comparison of best and worst shots. Journal of Sports Sciences 19(9), 727–733 (2001)
8. Kerick, S., et al.: The role of the left temporal region under the cognitive motor demands of shooting in skilled marksmen. Biological Psychology 58(3), 263–277 (2001)
9. Haufler, A., et al.: Neuro-cognitive activity during a self-paced visuospatial task: comparative EEG profiles in marksmen and novice shooters. Biological Psychology 53(2-3), 131–160 (2000)
10. Konttinen, N., Lyytinen, H., Viitasalo, J.: Preparatory heart rate patterns in competitive rifle shooting. Journal of Sports Sciences 16(3), 235–242 (1998)
11. Tremayne, P., Barry, R.J.: Elite pistol shooters: physiological patterning of best vs. worst shots. International Journal of Psychophysiology 41(1), 19–29 (2001)
12. Lacey, J.I.: Somatic response patterning and stress: Some revisions of activation theory. Psychological stress: Issues in research, 14–38 (1967)
13. Espinosa, P.D., et al.: Development of sensor-based measures of rifle marksmanship skill and performance (Deliverable to Advanced Brain Monitoring, Inc.). University of California, National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles (2008)
14. Rifle marksmanship (PCN 144 000091 00, MCRP 3-01A). U.S. Marine Corps (USMC), Albany, GA (2001)
15. Camm, A.J., et al.: Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. European Heart Journal (17), 354–381 (1996)
16. Nolan, R.P., et al.: Heart rate variability biofeedback as a behavioral neurocardiac intervention to enhance vagal heart rate control. American Heart Journal 149(6), 1137 (2005)
17. Green, S.C., Bavelier, D.: Effect of Action Video Games on the Spatial of Visuospatial Attention. Journal of Experimental Psychology 32(6), 1465–1478 (2006)
18. Cohen, H., et al.: Power Spectral Analysis of Heart Rate Variability in Psychiatry. Psychotherapy and Psychosomatics 68, 59–66 (1999)

Assessing Cognitive State with Multiple Physiological Measures: A Modular Approach

Lee W. Sciarini and Denise Nicholson

Institute for Simulation and Training
Applied Cognition and Training in Immersive Virtual Environments Laboratory
3100 Technology Parkway, Orlando, Florida, 32826
{Lsciarin,Dnichols}@ist.ucf.edu

Abstract. The purpose of this effort is to introduce a novel approach which can be used to determine how multiple minimally intrusive physiological sensors can be used together and validly applied to areas such as Augmented Cognition and Neuroergonomics. While researchers in these fields have established the utility of many physiological measures for informing when to adapt systems, the use of such measures together remains limited. Specifically, this effort will provide a contextual explanation of cognitive state, workload, and the measurement of both; provide a brief discussion on several relatively noninvasive physiological measures; explore what a modular cognitive state gauge should consist of; and finally, propose a framework based on the previous items that can be used to determine the interactions of the various measures in relation to the change of cognitive state.

Keywords: Augmented Cognition, Neuroergonomics, Physiological Measures.

1 Introduction

Advances in technologies, research, and interest in Augmented Cognition applications have all but guaranteed a future in which the physiological state of a human operator will impact the interactions with many, if not all, (closed-loop) systems. To the uninitiated, this statement almost assuredly conjures images of cyborgs and bionic beings that seemingly have given up their humanity. While the issue of being "more machine than man" may eventually become an ethical dilemma, current technology has not yet required its serious contemplation. Current technologies do, however, offer the opportunity to create systems in which the user is part of the interface. Through available technology, it is now possible to reexamine the human-centered system design (process) and include measurements of the human's state as a means to inform and even adapt the system.

Although researchers in fields such as Augmented Cognition (AUGCOG) and Neuroergonomics have begun to establish the utility of physiological measures for informing when to adapt systems, the use of such measures remains limited. While this may be partially explained by the high cost of equipment, it is more likely due to the lack of clear guidance for the use of multiple sensing devices to adapt systems.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 533–542, 2009.

This need was highlighted in 2007 by Reeves, Stanney, Axelsson, Young and Schmorrow [1] in their articulation of the near-, mid- and long-term goals of AUGCOG. The authors specifically noted that there were several impediments to the adoption of such technologies including: (1) the need for valid, reliable, and generalizable cognitive state gauges based on basic neurophysiological sensors; (2) real-time cognitive state classification based on basic cognitive psychology science and applied neuro-cognitive engineering; and (3) proof of effectiveness which demonstrates generalizable application of mitigations (i.e., the ability to control how/when mitigations are applied). Unfortunately, the ability to detect cognitive state through the use of various technologies based on different physiological indices currently poses problems. For example, sampling rates (i.e., resolution) of different measures may cause one technology to indicate a state change while another is reporting the previous state. Additionally, a particular measure may indicate the onset of a state change which may not be reported by other measures, and this dissonance may cause conflict when determining if an intervention is required.

As a tool, a "cognitive state gauge" is a vague concept which has the potential to include a wide range of contributing factors. When considering all of the possibilities, the goal of creating a valid and generalizable cognitive state gauge is a lofty one at best. In fact, the very idea of a cognitive state gauge poses issues of ambiguity similar to those of its conceptual springboard: *mental workload*. This vagueness, perhaps, is why such a measure has yet to be developed and/or proven effective in meeting the goals set by Reeves et al. [1].

Based on the multiple resource theory model [2],[3] and its idea that we draw from multiple distinct pools of cognitive resources, it is therefore proposed that instead of taking on the concept of a holistic cognitive state gauge, it is necessary to first manipulate specific cognitive resources and examine the physiological state as recorded by each of several synchronized measures. By using a modular approach which targets specific cognitive abilities in a controlled environment, it should be possible to build a reliable and generalizable cognitive state gauge based on basic cognitive psychology.

In order to describe a novel approach to assessing cognitive state with multiple physiological measures, this paper will provide a contextual explanation of cognitive state, workload, and the measurement of both. This will include a brief discussion of several relatively noninvasive physiological measures whose use, in concert, are proposed to present a solution to the impediments articulated by Reeves et al. [1]. Inspired by technologies described in the Augmented Cognition Technical Integration Experiment Report [4], the candidate measures that will be discussed include six non-cortical measures: eye blink rate (EBR); pupil dilation (PD); respiration rate (RR); heart rate (HR); heart rate variability (HRV); electrodermal response (EDR), and one cortical measure: electroencephalography (EEG). Specifically, this effort will explore what a modular cognitive state gauge (MCSG) should consist of and will also propose a framework. Additionally, a testbed based on the MCSG and proposed framework will be introduced for the purpose of determining the interactions of the various measures in relation to the change of cognitive state.

2 Measuring Cognitive State

It would be a futile effort to suggest that there is a way to measure cognitive state without first defining what is meant by the term cognitive state. For the purposes of this work, the idea that dynamic changes in human cognitive activity can be identified during task performance [5] allows us to define cognitive state as consisting of those aspects of cognitive ability which are called upon for the completion of a task.

While this may be an acceptable definition of cognitive state, it must be understood that there are numerous factors that contribute to cognitive state. For example, changing levels of fatigue or stress during task performance are responses, not indicators of the capacity of cognitive ability. Simply measuring the physiological response of fatigue and/or stress to a task would be to ignore the mechanisms that explain such responses. The mental capacity that allows for the successful completion of tasks should be the area of interest when investigating cognitive state. Ultimately, it is this capacity that, when taxed, results in performance decrements. The taxing of these mental capacities has been extensively investigated in various environments in order to understand the phenomenon of mental workload. If one intends to work toward the goals set forth by Reeves et al., it is necessary to understand what is meant by the term workload and to identify approaches that can be used for its measurement.

2.1 Workload

At the core, workload can be defined as the amount of demand(s) placed on an operator while attempting to accomplish something. Researchers have gone to great lengths to understand the effect of mental workload on performance. These researchers have proposed various theories and analogous models to explain how the human mind allocates its ability to handle information and task completion from the mundane to the complex. Byrne and Parasuraman [6] state that the general consensus on mental workload is based on theoretical models of resource and capacity for information processing. For this to be the case, it is accepted that humans have a finite amount of available cognitive resources which must be allocated and used to accomplish a task. In essence, mental workload is directly related to the proportion of the mental capacity an operator expends on the performance of a task [7],[8].

As a construct, workload is difficult to examine due to the seemingly limitless attributing variables. In his 2007 report to the Department of Transportation, Reinach [10] suggested that workload can be defined as the interaction between the demands of a task and an operator's ability to meet those demands. When considered in these terms, workload is viewed as being dependent upon an operator's level of training, expertise, experience, fatigue, stress, motivation, and his or her available cognitive abilities and resources for a given task. Of course, *task load* is an integral piece of the workload puzzle. Task load has been defined as the total amount of demands placed on an operator at a given moment in a situation [10]. For a contextual example, Hadley, Guttman, and Stringer [9] describe an air traffic controller's task load to include elements such as the volume and composition of traffic, routing complexity, and weather conditions. Therefore, in the context of this effort, workload is operationally defined as the demands on available cognitive ability and resources placed on an operator by the demands and complexity of a given task.

In 2002, Wickens provided a review of multiple resource theory (MRT) and its application with an updated four-dimensional model [3]. MRT suggests that there is not a single information processing source that can be tapped by an operator. Instead, in order to perform a task or tasks, Wickens [2],[3] proposes that an operator must draw from multiple distinct pools of resources simultaneously. Dependent upon the composition of the task(s), the operator may have to process information serially (if the task(s) require the same resource pool) or in parallel (if the task(s) require differing resource pools).

Central to this effort is the idea that Wickens' theory would view exceeding operator workload (resulting in a performance decrement) as a shortfall of available resources. Further, Wickens suggests that operators have a finite capability for information processing. In short, cognitive resources are limited and conflicts (operator overload) occur when an operator performs two or more tasks that require a single resource

Measuring Workload. Not surprisingly, numerous approaches for assessing workload have been developed, from relatively simple questionnaires to complex brain imaging techniques. Regardless of type, these approaches will generally fall into one of three distinct categories: performance, subjective, and physiological [8],[11],[12]. The following will discuss selected measures which are proposed for measuring cognitive state in this effort.

Performance Measures. As mentioned above, the measure of task performance is a widely used method of inferring the amount of workload experienced during the completion of a task. In general, research has shown that if performance is high (maintaining acceptable performance) then workload can be considered low. Conversely, low performance suggests high workload. However, there are various factors that contribute to the workload construct resulting in a non-linear relationship with performance. As a contributing factor to workload, performance does provide a quantifiable and potentially real-time (provided the parameters are known) method for assessing operator workload. The measurement of performance is generally separated into two main subcategories: primary and secondary task measures.

Primary Task Performance. On the surface, measuring primary task performance is a simple proposition. Unfortunately, this may not always be the case. Several factors can contribute to task difficulty experienced by an operator. For example, an increase in time pressure or the demands on cognitive resources may not always degrade performance [13]. The lack of performance decrement can be attributed to the operator's skill level or motivation to exert more effort to maintain an acceptable level of performance. These contributing factors can result in an incorrect assessment of operator workload due to the fact that acceptable performance is maintained while the operator is approaching the limitations of his or her cognitive capabilities.

Secondary Task Performance. The addition of a concurrently performed task to the primary task can be used to detect the workload of a primary task [14]. The goal of using a secondary task is to additionally tax the cognitive resources being used to complete the primary task. By doing so, an operator who is maintaining an acceptable level of performance is required to divert resources to the additional task and could potentially uncover his or her level of workload through an observable performance decrement in either the primary or secondary task. As suggested by multiple-resource theory [2][3], through the imposition of a secondary task that consumes the same resource(s) as the primary task, it should be possible to measure the excess resource(s) not utilized by the primary task.

Subjective Measures. One of the most commonly used methods for measuring workload is the NASA Task Load Index (TLX). The TLX is a subjective evaluation of workload that is completed by an operator upon completion of a task. The TLX is a multidimensional approach that measures workload by calculating a total workload score from six weighted subscales: mental demand, physical demand, temporal demands, performance, effort, and frustration level. These six subscales are based on extensive research and psychometric analyses from a wide range of contexts [15].

While asking an operator to evaluate his or her own level of workload following completion of a task has utility, most tasks are not static, isolated events and post hoc assessment, by its nature, would fail to offer real-time assistance to the operator. People are expected to perform in complex and dynamic environments which tend to evolve over time with the emergence of information. The complexity and propensity for real world operations to present novel and often hard-to-predict situations makes real time and predictive state assessment extremely intriguing as a way to inform potential mitigations to operator workload. While subjective ratings such as the TLX are useful for eliciting overall task workload assessment, they lack the ability to provide real-time assessment without intrusion.

Physiological Measures. The idea that physiological measures may assess workload is not a new one. For example, in their report for NASA, Scerbo, Parasuraman, Di Nocero, and Prinzell discussed the efficacy of using physiological measures for adaptive automation [16]. Their effort highlighted four promising physiological measures that could be used to assess mental workload: eye blink, respiration, cardiovascular activity, and speech measures. Additionally, EEG was discussed as a cortical measure that may inform the adaptation of automation.

It should come as no surprise that there are numerous methods that use physiological measurement technologies to assess cognitive state. Each of these methods use a unique approach to their measurement and assessment, a detail that must be addressed. The argument that one measure is adequate for operational systems will not suffice in the face of multidimensional tasks which are carried out in dynamic environments. Although, the use of multiple measures, as stated previously, presents confounding factors which must be considered. The responsiveness of one measure to the change of an operator’s state may not occur within the same time frame as another measure. One measure may provide a global view of operator state while another may be better suited to detect subtle changes based on discrete events and/or situations. Confusion and even catastrophe can occur if system(s) dependent on these differing physiological measures are based on conflicting indications of operator state change. In order to achieve the goal of assessing cognitive state through the use of multiple physiological measures, it is important to discuss candidate physiological measures. These measures include six non-cortical measures: EBR; PD; RR; HR; HRV; EDR and one cortical measure: EEG. Table 1 provides an overview of each candidate technology.

Table 1. Overview of candidate physiological measures

Type	Description
EBR	Shown to be a useful measure of mental workload [17],[18]. Several laboratory and field studies have shown that blink rate decreases with an increase in task difficulty (e.g., [19],[20],[21],[22])
PD	Shown to decrease or increase depending on autonomic response. Pupil dilation is an important measure of mental workload [7] and has been used numerous times as a global measure of workload. Increased pupil diameter has been observed with an increase in resource taxation [22]
RR	Proposed as a useful physiological indicator of the state of an operator. Increased respiration rate along with a decrease in the depth of inspiration have been associated with increases in stress and cognitive demand
HR	Likely candidate for measuring cognitive workload. Wilson & Eggemeier [23] suggest that heart rate could predict and be an overall indicator of workload. This is supported by a series of workload studies showing that heart rate was the most favorable physiological measure
HRV	Decreases with the increase in heart rate. An increase in workload results in a decrease in heart rate variability [26] when compared to the rest state [8]. Of particular interest for the measurement of mental effort is the varying duration of time between heartbeats, the inter-beat interval [27]
EDR	Measurable change of electrical activity of the skin as a result of sweat gland activity capable of indicating stress-strain, emotion, and arousal [28]. One of the several measures of EDR, Skin Conductance Level (SCL) is measured by the application of a constant voltage to the skin via electrodes in order to measure conductance. Research has shown that there can be a significant increase of SCL across workload conditions [29]
EEG	Provides the total amount of the electrical brain activity of active neurons that can be recorded on the scalp through the use of electrodes [30]. Berka et al. validated use of EEG for measuring task engagement and mental workload. An investigation utilizing their task engagement and mental workload measures had promising results showing that participants' EEG-workload index increased on tasks with increasing difficulty and working memory load. Similarly, EEG-engagement was shown to be related to the processes required for completing vigilance tasks [30]

3 Modular Cognitive State Gauge (MCSG)

As stated in the introduction, the objective of this work is to define a useful approach for using multiple physiological measures to assess one's cognitive state. The paradigm presented here aims to segregate specific contributors to mental workload for measurement. It is proposed that by systematically exploring the manner in which each physiological measure correlates to performance and to each other in targeted areas, a cognitive state gauge that meets the validity, reliability, and generalizability requirements set forth by Reeves et al. [1] can be created.

It is proposed here to use Wicken's MRT model [2],[3] as a practical guide for investigating multiple physiological sensors and their combined ability to predict performance decrements in specific cognitive resource areas. After compiling an understanding of what to expect for particular cognitive resources through empirical research, the MCSG should begin to take shape (Figure 1). Essentially, it is proposed that by parsing out individual cognitive resources (e.g., visual, auditory, spatial, etc.) into modules, they can be empirically investigated and then integrated into a generalizable cognitive state gauge.

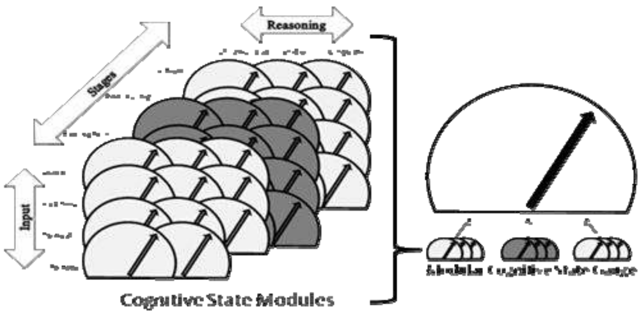


Fig. 1. MRT-based cognitive modules towards a modular cognitive state gauge

By using this modular approach, potential issues with the use of multiple sensors can be identified and addressed as the modules are investigated. For example, HRV and EEG, as discussed previously, have both been shown to be useful for measuring workload. Interestingly, Gohara et al. [31] discovered that HRV becomes less sensitive when measured during a state of fatigue. Discrepancies between measures like these could present serious consequences to the accuracy of any cognitive state gauge if the input were not understood.

While it may seem daunting to examine the multitude of cognitive resources in such a systematic way, the great potential of previous efforts conducted by various academic, private, and government institutions [1],[32] will undoubtedly contribute to the compilation of the proposed MCSG. Of course, once a sufficient amount of modules are understood, the next challenge will be integrating them into a unified gauge. There could be a variety of approaches to accomplishing this task and these will undoubtedly be discussed in subsequent investigations.

4 Proposed Implementation

While the investigation of each module may be unique, the following should provide at least the basic heuristics to determine a course of action. It is proposed here to identify experimental methods from previous foundational studies which focus on the cognitive resource of interest and adopt those efforts for investigation with multiple physiological sensors. Once an effort has been identified, it is suggested that the three types of workload measures described in section 2.1.2 (performance, subjective, and physiological) should be collected for the new investigation. By following this implementation, it is assumed that any new confounds should be limited to the new measure(s).

When determining which physiological measures to use, the most relevant devices should be considered first. For example, eye tracking would be an obvious choice for an investigation exploring visual search and attention but may not provide meaningful data for an effort solely focused on the area of auditory attention.

Once the physiological sensors have been determined it is recommended that all experimental components are synchronized. While it would be inappropriate, and in some cases impossible, to attempt to force physiological measuring devices into having identical sampling rates, they can be synchronized to each other and the experimental environment. At a minimum, a timestamp indicating the beginning and conclusion of an experimental trial common to all data logs should be the included. Additionally, synchronously recording performance in the experimental environment with the selected physiological measures will allow for successful matching of changes in performance for observation. For example, in an effort to use multiple sensors for an adaptive learning system, Vartak et al. [33] proposed a block processing model in order to synchronize and evaluate the volumes of physiological data from multiple measures. Using an approach similar to the one found in Varatak et al.'s model should prove to help streamline the data collection and perhaps even aid in the development of future AUGCOG applications. Finally, perceived levels of workload can only be obtained by asking. Collecting subjective measures, while not dynamic, can be extremely useful in providing consistency across participants.

5 Future Work and Conclusion

Previous research using a dynamic spatial task showed that highly skilled participants outperformed those with lower skills when evaluated on spatial ability tests [34]. Using a similar task and methods, we will investigate the modular approach described here through the implementation outlined in section 4.

This paper proposed an approach that can be used determine under what conditions multiple minimally intrusive physiological sensors can be used together and validly applied to a cognitive state gauge. Through the use of the model and implementation proposed, we are confident that various physiological measures can be used to accurately measure changes in cognitive state while meeting the goals set forth by Reeves et al. [1].

References

1. Reeves, L., Stanney, K., Axelsson, P., Young, P., Schmorrow, D.: Near-term, Mid-term, and Long-term Research Objectives for Augmented Cognition (a) Robust Controller Technology and (b) Mitigation Strategies. In: 4th International Conference on Augmented Cognition, pp. 282–289. Strategic Analysis, Inc., Arlington (2007)
2. Wickens, C.D.: Processing Resources in Attention. In: Parasuraman, R., Davies, D.R. (eds.) Varieties of Attention, pp. 63–102. Academic Press, London (1984)
3. Wickens, C.D.: Multiple Resources and Performance Prediction. Theoretical Issues in Ergonomics Science 3, 150–177 (2002)

4. St John, M., Kobus, D.A., Morrison, J.G.: DARPA Augmented Cognition Technical Integration Experiment (TIE) Tech. Report 1905. United States Navy SPAWAR Systems Center, San Diego (2003)
5. St. John, M., Kobus, D.A., Morrison, J.G., Schmorrow, D.: Overview of the DARPA Augmented Cognition Technical Integration Experiment. Int. J. of HCI 17, 131–149 (2004)
6. Byrne, E.A., Parasuraman, R.: Psychophysiology and Adaptive Automation. Biological Psychology 42, 249–268 (1996)
7. Kahneman, D.: Attention and Effort. Prentice-Hall, Englewood Cliffs (1973)
8. Brookhuis, K.A.: Psychophysiological Methods. In: Stanton, N., Hedge, A., Brookhuis, K., Salas, E., Hendrick, H. (eds.) Handbook of Human Factors and Ergonomics Methods, pp. 17-1–17-5. CRC Press, Boca Raton (2005)
9. Hadley, G., Guttman, J., Stringer, P.: Air traffic control specialist performance measurement database. (Tech. Report DOT/FAA/CT-TN99/17). Federal Aviation Administration William J. Hughes Technical Center, Atlantic City (1999)
10. Reinach, S.: Preliminary Development of a Railroad Dispatcher Taskload Assessment Tool: Identification of Dispatcher Tasks and Data Collection Methods. Tech. Report DOT/FRA/ORD-07/13. U.S. Department of Transportation Federal Railroad Administration Office of Research and Development, Washington (2007)
11. O'Donnell, R.D., Eggemeier, F.T.: Workload Assessment Methodology. In: Boff, K.R., Kaufman, L., Thomas, J.P. (eds.) Handbook of Perception and Human Performance. Cognitive Processes and Performance, vol. II, pp. 42-1–42-49. Wiley, New York (1986)
12. Wierwille, W.W., Eggemeier, F.T.: Recommendations for Mental Workload Measurements in a Test and Evaluation Environment. Human Factors 35, 263–281 (1993)
13. Wickens, C.D.: Workload and Situation Awareness. In: Hancock, P., Desmond, P. (eds.) Stress, Workload and Fatigue, pp. 443–450. Lawrence Erlbaum Associates, Mahwah (2001)
14. Paas, F., Juhani, E.T., Tabbers, H., Van Gerven, P.W.M.: Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. Educational Psychologist 38, 63–71 (2003)
15. Hart, S.G., Staveland, L.E.: Development of the NASA Task Load Index (TLX): Results of Experimental and Theoretical Research. In: Hancock, P., Meshkati, N. (eds.) Human Workload, pp. 138–183. North-Holland, Amsterdam (1988)
16. Scerbo, M.W., Freeman, F.G., Mikula, P.J., Parasuraman, R., Di Nocero, F., Prinzel III, L.J.: The Efficacy of Psychophysiological Measures for Implementing Adaptive Technology. Tech. Report NASA/TP-2001-211018. NASA Langley Research Center, Hampton (2001)
17. Morris, T.L., Miller, J.C.: Electooculographic and Performance Indices of Fatigue During Simulated Flight. Biological Psychology 42, 343–360 (1996)
18. Veltman, J.A., Gaillard, A.W.K.: Physiological Indices of Workload in a Simulated Flight Task. Biological Psychology 42, 323–342 (1996)
19. Backs, R.W., Ryan, A.M., Wilson, G.F.: Psychophysiological Measures During Continuous and Manual Performance. Human Factors 36, 514–531 (1994)
20. Hankins, T.C., Wilson, G.F.: A Comparison of Heart Rate, Eye Activity, EEG and Subjective Measures of Pilot Mental Workload During Flight. Aviation, Space & Environmental Medicine 69, 360–367 (1998)
21. Wilson, G.F.: An Analysis of Mental Workload in Pilots During Flight using Multiple Psychophysiological Measures. Int. Jnl. of Aviation Psych. 12, 3–18 (2002)
22. Beatty, J.: Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources. Psychological Bulletin 91, 276–292 (1982)

23. Wilson, G.F., Eggemeier, F.T.: Psychophysiological Assessment of Workload in Multi-task Environments. In: Damos, D.L. (ed.) Multiple-task Performance, pp. 329–360. Taylor & Francis, London (1991)
24. Roscoe, A.H.: Heart Rate as Psychophysiological Measure for Inflight Workload Assessment. Ergonomics 36, 1055–1062 (1993)
25. Wilson, G.F.: Applied Use of Cardiac and Respiration Measures: Practical Considerations and Precautions. Biological Psychology 34, 163–178 (1992)
26. Mulder, L.J.M., De Waard, D., Brookhuis, K.A.: Estimating Mental Effort Using Heart Rate and Heart Rate Variability. In: Stanton, N., Hedge, A., Brookhuis, K., Salas, E., Hendrick, H. (eds.) Handbook of Human Factors and Ergonomics Methods, pp. 20-1–20-8. CRC Press, Boca Raton (2005)
27. Boucsein, W.: Electrodermal Measurement. In: Stanton, N., Hedge, A., Brookhuis, K., Salas, E., Hendrick, H. (eds.) Handbook of Human Factors and Ergonomics Methods, pp. 18-1–18-12. CRC Press, Boca Raton (2005)
28. Shi, Y., Ruiz, N., Taib, R., Choi, E.H.C., Chen, F.: Galvanic Skin Response (GSR) as an Index of Cognitive Load. In: CHI 2007, San Francisco (2007)
29. Akerstedt, T.: Ambulatory EEG Methods and Sleepiness. In: Stanton, N., Hedge, A., Brookhuis, K., Salas, E., Hendrick, H. (eds.) Handbook of Human Factors and Ergonomics Methods, pp. 21-1–21-7. CRC Press, Boca Raton (2005)
30. Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P.D., Craven, P.L.: EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks. Aviation Space and Environmental Med. 78(Suppl. 5), B231–B244 (2007)
31. Gohara, T., Mizuta, H., Takeuchi, I., Tsuda, O., Yana, K., Yanai, T., Yamamoto, Y., Kishi, N.: Heart Rate Variability Change Induced by the Mental Stress: The Effect of Accumulated Fatigue. In: 15th Southern Biomed. Eng. Conf., pp. 367–369 (1996)
32. Reeves, L., Schomorrow, D.: Augmented cognition foundations and future directions - Enabling "anyone, anytime, anywhere" applications, pp. 263–272. Springer, Berlin (2007)
33. Vartak, A., Fidopiastis, C., Nicholson, D., Mikhael, W., Schmorrow, D.: Cognitive State Estimation for Adaptive Learning Systems using Wearable Physiological Sensors. In: 1st Intl Conf. on Biomedical Electronics and Devices (2008)
34. Sims, V., Mayer, R.: Domain Specificity of Spatial Expertise: The Case of Video Game Players. Applied Cognitive Psychology 16, 97–115 (2002)

Neuro-NIRS: Analysis of Neural Activities Using NIRS

Hiroshi Tamura[1], Miki Fuchigami[2], and Akira Okada[2]

[1] Institute for HUMAN INTERFACE, C1803 Portpia plaza, 6-14 Minatojima nakamachi, Chuoku, Kobe City, JAPAN 650-0046
[2] Osaka City University, Sugimotocho, Osaka City, Japan
tamura@mobilergo.com

Abstract. Analysis method for quick component of NIRS is explained. In order to compare quantitatively the amount of quick component, absolute average operation is applied. The result Q is then formulated as the product of magnitude M and density D. By characteristic differences of M and D functional features of each channel are discussed. The method is applied to text entry task to mobile phone. Most of 34 channels under study D values diverged in rest state, but they converged at task state. 6 channels among 34, showed specific responses to specific task.

Keywords: neural activities, NIRS, quick components, text entry, mobile phone.

1 Introduction

Fruitful sessions on NIRS (near infrared spectroscopy) were organized by Monica Fabiani at HCII 2005 held in Las Vegas. By that time, two Japanese companies have developed general purpose NIRS machines, which are capable of simultaneous measuring of oxy and deoxy hemoglobin concentration in the brain surface tissues. Since NIRS has been used in the field of hemo-dynamic studies until then, the new machines are still used mainly to trace oxy hemoglobin and, so far deoxy hemoglobin is often disregarded. Tamura [1] emphasized need to develop analysis methods considering oxy and deoxy hemoglobin at the same time. In HCII 2007 a method to analyze NIRS data in the two dimensional coordinate plane, taking the total (oxy + deoxy) and difference (oxy - deoxy) variables to the axes. The paper showed the total component consisted of, mainly, slow component, while the difference component consisted of slow and quick components.

The slow and the quick components are separated by applying running average of 2 seconds to the difference component. Slow component normally has continuous waveform, while the quick shows the form of pulse train of variable height. The interests of hemodynamic studies have been in the slow component and the quick component is regarded as noise, and no attentions have been paid to the quick component. Although hemodynamic studies are not interested in quick component, they already know the existence of quick signals and regard it fatal for successful NIRS measurements. Many people tried to decrease noise by taking care of external factors like optical contacts to the skin. But the quick component is large in amplitude, often much larger than the slow. It is not reasonable to regard it noise.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 543–552, 2009.

We are convinced that the quick component is not noise but essential physiological signal. We have made measurement [2] of quick component distributions using 34 to 48 channels NIRS, provided by Shimadzu and applied it to various tasks. It was confirmed that the quick distributions changes according to the task, such as text input, hand up down, finger sign and button selection, etc. Slow component are not stable and effected by various factors, but quick components distributions are stable, although they are varying by conditions.

In this paper, the author presents the model of quick and slow component generation with regard to blood circulation systems. Secondly analysis method to draw neural activities included in each NIRS signals is explained. The quick signal from a single channel of NIRS is results of summative activities of small brain area, and speculative model of activities is introduced in this paper.

2 Quick Components

2.1 NIRS Coordinates

Fig. 1 is showing Oxy, Deoxy, and Total Hb (hemoglobin) recording of 6 channels from NIRS. The upper traces correspond to oxy, the lower to deoxy, and the middle to total Hb. The channel number is indicated at the middle of the left side of each chart, Looking to the channel 26, oxy and deoxy Hb contains big high frequency components, which we call quick components, while total Hb contains less. The traces of 29 and 32 contain less quick components, in both oxy and deoxy, as well as total Hb.

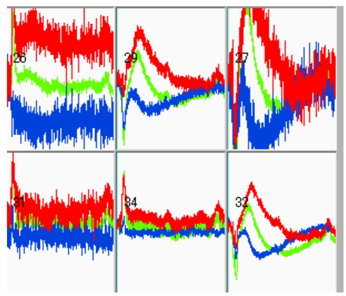


Fig. 1. Raw traces of NIRS

The NIRS variables, oxy and deoxy Hb can be replaced by total and difference Hb as below:

$$\text{Total Hb:} \quad \text{Total} = \text{Oxy} + \text{Deoxy} \quad (1)$$

$$\text{Difference Hb:} \quad \text{O\_D} = \text{Oxy} - \text{Deoxy} \quad (2).$$

Fig. 2 is showing traces of above variables. The quick components of oxy Hb and deoxy Hb are in opposite phases, thus by the addition of quick components, they are almost compensated, while by subtraction, the quick component is doubled in amplitude.

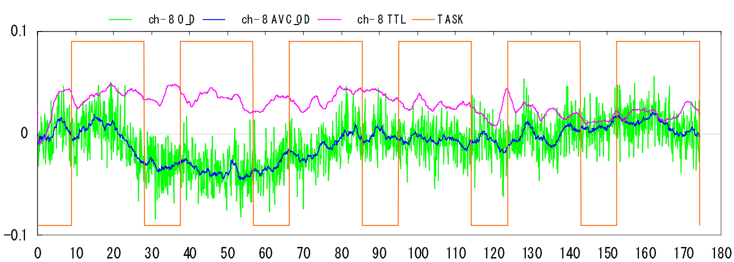


Fig. 2. Total Hb(red) and difference Hb(green) and the slow O_D (blue) among the middle of difference(O_D) trace. Square trace indicates on and off of task.

Applying moving average of 2 second, quick and slow components of Total and O_D can be derived.

Slow total Hb:	slow total=Average (Total)	(3)
Slow difference Hb:	slow O_D= Average (O_D)	(4)
Quick total Hb:	Quick t= Total − Average(Total)	(5)
Quick difference Hb:	Quick O_D = O_D − Average(O_D)	(6)

Thus the analysis of quick components of NIRS can be concentrated to the quick O_D. The quick components derived from eq. (5) and (6) are shown in Fig. 3. The magnitude of quick total is about 1/10 of quick O_D, the characteristics of the latter are mainly discussed below.

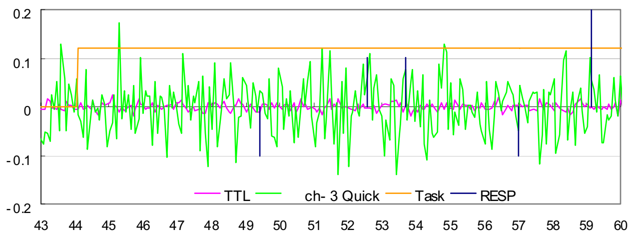


Fig. 3. Traces of Quick O_D and Quick total

2.2 Analysis of Quick Components

For the comparison of quick components, derived from various channels at various tasks, unified quantitative formulation of quick component is necessary. An appropriate function applicable to noise like signal as quick components is Avedev (Microsoft excel function), which is average of absolute difference of data and their mean value. In our analysis, slow component is subtracted from O_D signal, thus the mean value of quick component is assumed to be zero. In short, avedev can be called the time average of absolute value of quick.

In the previous papers [2], long term (>30second) distributions of quick components around different channels and their changes are reported. A distribution proper to the case of this paper is shown in Fig. 9 below.

Avedev is stable in value, when time length is taken 10 seconds or longer. It change considerably as averaging time is shorten, reflecting change of brain activities under the recording. In this paper, a method to evaluate short term change in brain activities is considered by introducing two parameters, M (magnitude)and D (density), related to Q.(avedev of quick component). The relation of these parameters is formulated as below.

$$Q = M * D \tag{7}$$

Here

$$M = \max(\max(\text{member}), -\min(\text{member})) \tag{8}$$

member: data within time span of averaging, and the number of data is denoted by N.

Following the above formulation, relation below is derived.

$$0 < 1/N < D < 1 \tag{9}$$

Inequality (9) states that the density remains with in the bound of 0 and 1. Further studies are concentrated to the change of M and D, in succession of averaging cycles, which are taken 1, 2 and 5 second, for comparison, depending on the tasks.

Imagine pulse trains of full, half, quarter, etc density as shown in Fig 4a. Relations of Q and M are shown by the lines of positive slope in Fig.4b. D is larger at the upper side of the slope lines. When quick component data are plotted on Q and M plane, scatter graph as in Fig. 4c is obtained. The data points are scattered around a positive slope line, and the slope is definite to member size, independent of channel. The member size increases in accordance to averaging interval of avedev (Q). Thus the inclination of slope decreases with increase of averaging interval. The degree of scatter is different depending on channel.

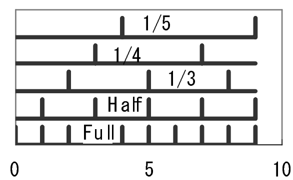


Fig. 4a. Pulse trains

Generally, the channels with large M show big scattering. The slope changes the inclination also depending on the tasks.

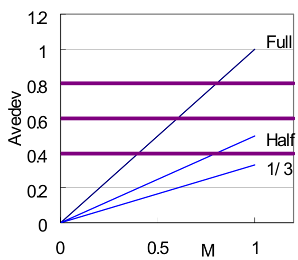


Fig. 4b. Theoretical relationsof Q and M

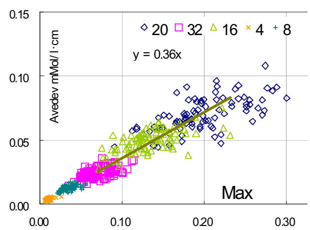


Fig. 4c. Actual distribution of Q and M

Now the scatter plot of NIRS data is considered on density (D) and magnitude (M) plane (Fig. 4d).

The positive slope lines in Fig. 4b are mapped to separate horizontal lines in Fig. 4d. Next the mapping of horizontal lines (Q = constant) in Fig. 4b to Fig. 4d is considered.

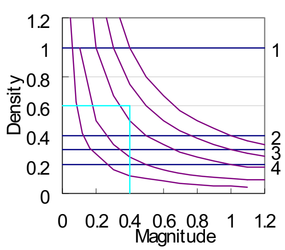


Fig. 4d. Relation of D and M

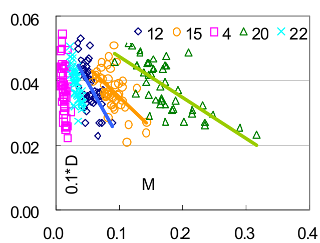


Fig. 4e. Data plot in D and M

They are transformed to hyperbola in Fig. 4d. The scatter graph Fig. 4c is transformed to Fig. 4e. The ordinate is extended in Fig. 4e. The data area of Fig. 4e is the square zone between origin and point 0.4, 0.6) of Fig. 4d.

When D-M data are plotted for one task period, and adjacent dots are connected by lines, trajectory as in Fig. 4f are obtained. The channel 20, shown in green triangles, include active change of M in short interval of time. On the contrary, channel activities are mainly in D change for channel 27 and 12.

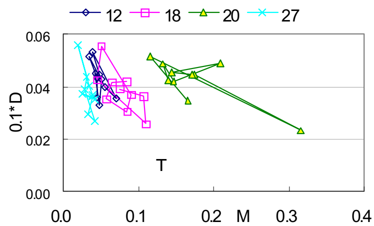


Fig. 4f. D-M trajectory for a Task period. 1019 1454 HIGHX.

3 Sources of Slow and Quick Components

3.1 Hemodynamic Systems

The major NIRS signals, oxy and deoxy Hb, have each the different source. The source of oxy Hb supply is in lung oxygen exchange. Oxy Hb travels a long way up to brain passing through various barriers. Thus the hemodynamic responses are normally slow. The source of deoxy Hb is neural activities in the brain, which is quick in nature. The quick component is gradually smoothed as it goes through cilliary vessel, and finally changes into slow deoxy Hb component. Thus the quick component directly recorded by NIRS is most likely reflecting neural activities of brain.Fig.5 is showing the model of hemodynamic systems related to NIRS.

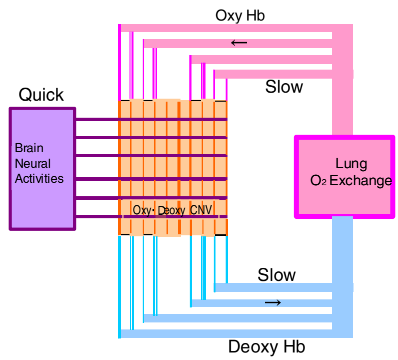
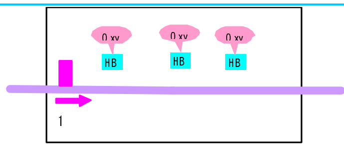


Fig. 5. Hemodynamic model of O_2 circulation

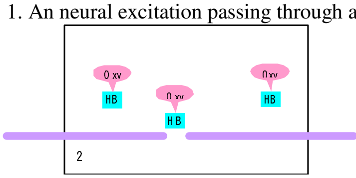
3.2 Oxy/ Deoxy Conversion in Neural Net

Fig. 6 explains the process converting oxyHb into deoxy Hb after arrival of an neural excitation using the quantum model. After the conversion process one unit of oxyHb decreased and one unit of deoxyHb increased. Resultantly, double unit of deoxyHb increase is observed by one neural excitation. The model is in accordance to the quick component increase of O_D.

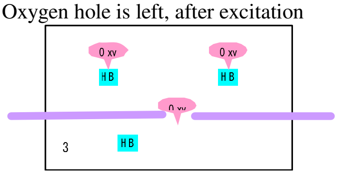
Neural excitation is energy consuming process. The neural system has to supply energy by support of ATP cycle immediately after excitations. In the recovery process after excitation, certain autofluorescence is reported [8]. One possibility of quick optical activities sensed by NIRS is after effect of neural excitations.



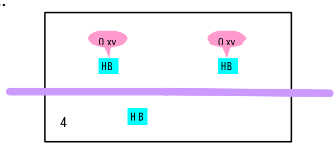
1. An neural excitation passing through axon



2. Oxygen hole is left, after excitation



3. oxyHb supplies oxygen to neuron, Hb is released.



4. local population after recovery of excitation.

Fig. 6. Quantum model of deoxidization of oxyHb after neural excitation

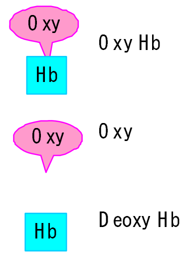


Fig. 7. Notations of O_2 carriers

Two types of general purpose NIRS machine are in market. One provided by Hitachi makes use of continuous modulation of infra red light, while the other provided by Shimadzu, makes use of sampling method. The latter looked annoying at first glance, because of prominent noise like responses. Now we convinced that they are providing significant physiological data. In case of machine using continuous modulation light, in order to cut modulation signal component at output stage, strong high cut filter is necessary. The filter might cut quick component considerably. But careful analyzer might detect quick component as well.

4 Text Entry Task

Text entry task to mobile phone is analyzed in this chapter. The user participated in experiments were female students of university, informed consented in written form. The students are accustomed to enter in multi-tap telephone key entry, and predictive Kanji conversion method.

The subject sat in front of CRT, where text of up to 20 characters are shown on the task phase, and on rest phase, colored picture of nature are shown. Subject held the mobile phone even in rest phase and stayed ready to start. Task and rest phases are 20 and 10 seconds, and the task rest cycle repeated 6 times. Types of text are different cycle by cycle, so that mental effort might be different at each cycle.

4.1 Task Description

Text types of 6 task cycles are listed below:

1. iroha	いろはにほへと	Japanese alphabet like ABC
2. reverse iroha	とへほにはろい	reverse order (meaningless) text
3. Kanji conversion	色は匂へど 散りぬるを	classic sentence, not fully fit to
4. kanji conversion	我が世誰ぞ 常ならむ	predictive conversion
5. number entry	119999965566666	same key location as task 1
6. number entry		same key location as task 2

Task 1 is starting part of Japanese alphabet like ABC, which has deep meaning. Task 2 is to enter reverse order text which has no meaning. Text entry speed is not different whether the text is with or without meaning. Task 3 and 4 are text entry with predictive Kanji conversion. But the text are somewhat different from present-day lettering, subject has to think about to overcome the difference. Task 5 and 6 are number entry, the key location and the number of key presses is equal to task 1 and 2. Conversion of letter to multi-tap operation is not required.

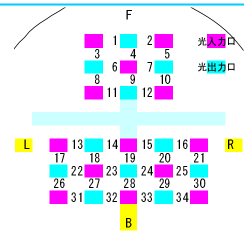


Fig. 8. Optode arrangement (34ch)

4.2 Optode Arrangement

Optode arrangement for present experiment is shown in Fig.8. The red optode indicate light emitting and blue light sink. The number between optodes indicates channel number. The sampling frequency for the arrangement was 100ms. Trace of O_D and total Hb for channel 8 is shown in Fig. 2 together with task/ rest signal.

The quick component recorded in each channel is compared by taking avedev channel by channel , averaging time 180sec.

The result is shown in Fig.9. The highest beak is observed at channel 20, and a sub peak at channel 16. At the 5 channels in the front most, quick components were low.

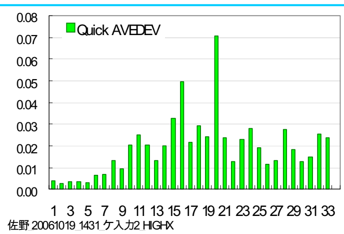


Fig. 9. Quick component distribution around the channels (180sec., ordinate in mMol/lcm)

4.3 Time Course of Magnitude and Density

Quick component are clustered by 2 second, and 2 second avedev was applied for each data. Then M by using eq.(8), and D by using eq.(7) is determined to each cluster.

M and D change their value by time. Because D is 10 times larger in value, in drawing graph, 0.1D is adopted. Fig 10 is showing 0.1D, M, Q and task on/off signal in time. It is interesting, M and Q changes considerable by channel and task, 0.1D is changing in definite range.

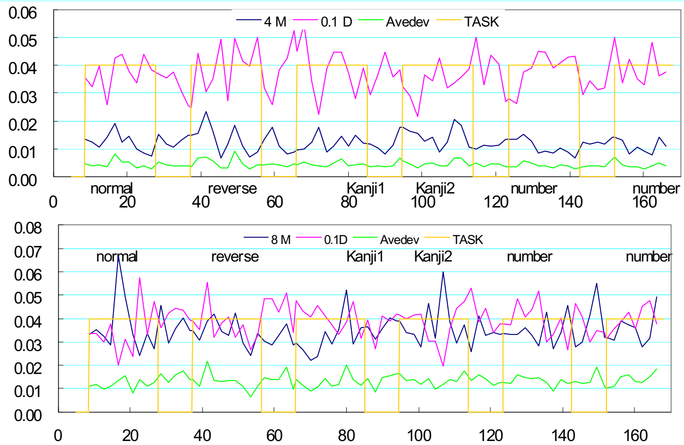


Fig. 10. Change of cluster parameters Q, M, and 0.1*D in time sequence

The variations of D and M are showing the neural processing in task and rest intervals in the channels.

5 Density Relations in Task and Rest Interval

For simplicity of discussion, average of 0.1D values of all the channels is compared in the course of 6 tasks/rest intervals (Fig. 11). Interesting finding is that density of many channels diverges (increase or decrease) at rest, while it converges at task state. Most of the channels show the feature.

In order to confirm the relations, density of all the 34 channels is averaged, and standard deviation of densities is calculated at each task and rest state.

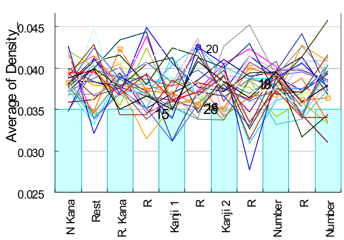


Fig. 11. Density variation at task and rest, comparison of 34 channels 1019 1454 HIGHX

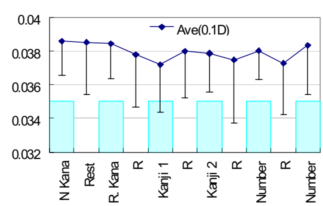


Fig. 12. Average and standard deviation density at task and rest

The results are shown in Fig. 12. The blue line shows the average and the ladder extended from the average is indicating the standard deviation. The standard deviation is lager at rest than at task, at most of the intervals.

6 Density Characteristics at Different Locations

Different locations of brain might have different density characteristics. As shown in Fig. 8, 34 channels are located in three parts of brain. One is in frontal, the second in left apical, and the last in right apical area. Density characteristics with task/rest interval for the channels in each area are illustrated in Fig. 13. The channel location is illustrated by the location maps associated to density graph.

In frontal area, density of most of the channels increases at rest, except for channel 7 and 12. In the apical area also, most channels are density increasing at rest. Some channels which are density increasing at task, 18and 27in left, and 20, 28, 33 in right apical area are accumulated Fig. 14. Each channel seems to have specific task at which it might show high density.

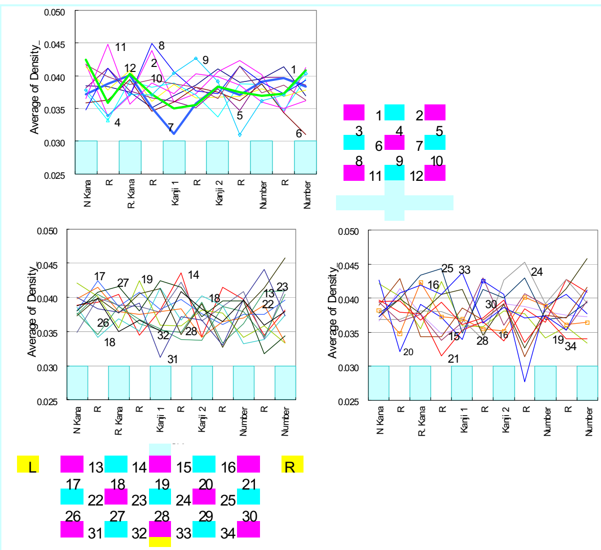


Fig. 13. Average of density within task/rest for all channels

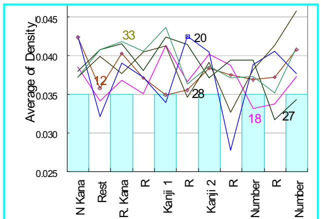


Fig. 14. Density increasing at task

7 Conclusion

Extended analysis method for NIRS quick component is presented. Quick component analysis is going to be most promising to study working brain.

Finally, influences of quick brain activities are partly represented to NIRS trajectories which are derived from slow components. By comparing NIRS trajectories, people might realize, there are channels which increase in oxy or total hemoglobin but less in neural activities. On the contrary, in some channels total or oxy Hb remains steady while the channels are busily functioning,

Hemodynamic studies expect the working brain needs more oxygen supply. Thus the areas supplied with rich oxygen are tentatively working. But this is only the expectation, but not evidence. We have to find evidences related to parts of the brain actually functioning.

Hemodynamic and neural functions are both significant aspects of brain studies. Cerebral infarction might cause hemodynamic problems, which might result in neural dysfunctions. Rehabilitation is the process of recovering hemodynamics and neural function. Neuro-NIRS is a new significant aspect of NIRS studies.

References

[1] Tamura, H., Omori, M., Choui, M.: NIRS trajectories in Oxy-Deoxy Hb plane and the trajectory map to understand brain activities related to human interface. In: Smith, M.J., Salvendy, G. (eds.) HCII 2007. LNCS, vol. 4557, pp. 994–1003. Springer, Heidelberg (2007)

[2] Tamura, H., Fuchigami, M., Okada, A.: On Physiological Roles of Quick Components in Near Infrared Spectroscopy (in Japanese). In: Proc. Mobile Interactions 2008, Society of Mobile Interactions (2008)

[3] Fuchigami, M., Okada, A., Tamura, H., Omori, M.: A Comparative Study of Brain Activities Engaged in Interface Operations by Means of NIRS Trajectory Map. In: Smith, M.J., Salvendy, G. (eds.) HCII 2007. LNCS, vol. 4557, pp. 830–839. Springer, Heidelberg (2007)

[4] Tamura, H., Omori, M., Choui, M.: NIRS trajectories in Oxy-Deoxy Hb plane and the trajectory map to understand brain activities related to human interface. In: Smith, M.J., Salvendy, G. (eds.) HCII 2007. LNCS, vol. 4557, pp. 994–1003. Springer, Heidelberg (2007)

[5] Tamura, H., Fuchigami, M., Okada, A.: Analysis of Text Entry behavior by NIRS Quick Components. In: Proc. Mobile Interactions 2009, pp. 109–114. Society of Mobile Interactions (2009)

[6] Fuchigami, M., Okada, A., Tamura, H.: Analysis of NIRS Quick Components in direction-selective button operations. In: Proc. Mobile Interactions 2009, pp. 115–120. Society of Mobile Interactions (2009)

[7] Fuchigami, M., Tamura, H., Okada, A.: Brain Activities supporting Finger Operations. In: Neuro-NIRS, HCII 2009

[8] Shibuki, et al.: Dynamic imaging of somatosensory cortical activity in the rat visualized by flavoprotein autofluorescence. Journal of Physiology 549, 919–927 (2003)

Eye Movements and Pupil Size Reveal Deception in Computer Administered Questionnaires

Andrea K. Webb[1], Douglas J. Hacker[1], Dahvyn Osher[1], Anne E. Cook[1], Dan J. Woltz[1], Sean Kristjansson[2], and John C. Kircher[1]

[1] University of Utah, 1705 E. Campus Center Dr. Rm 327, Salt Lake City, UT 84112
[2] Washington University, St. Louis, MO
andreakwebb@gmail.com

Abstract. An oculomotor test is described that uses pupil diameter and eye movements during reading to detect deception. Forty participants read and responded to statements on a computerized questionnaire about their possible involvement in one of two mock crimes. Twenty guilty participants committed one of two mock crimes, and 20 innocent participants committed no crime. Guilty participants demonstrated speeded and accurate reading when they encountered statements about their crime and increases in pupil size. A discriminant function of oculomotor measures successfully discriminated between guilty and innocent participants and between the two groups of guilty participants. Results suggest that oculomotor tests may be of value for pre-employment and security screening applications.

Keywords: Oculomotor measures, pupil size, deception.

1 Introduction

Many government agencies routinely conduct credibility assessments to screen applicants for positions in intelligence, security, law enforcement, immigration, and public transportation. Errors in classifying an individual as truthful or deceptive in these settings can have serious consequences for the individual and society. Current screening techniques rely primarily on the polygraph; however, other techniques have been used, such as self-report measures of integrity or personality, behavioral analyses, or speech content analyses [1-3]. Recently, questions have been raised about the validity of the polygraph for screening and its susceptibility to countermeasures. For example, the National Research Council (NRC) was critical of the polygraph for pre-employment screening and highlighted the need for "an expanded research effort directed at methods for detecting and deterring major security threats, including efforts to improve techniques for security screening..." ([4], p. 8). Similarly, self-report integrity tests for screening potential employees have been criticized due to questions about their effectiveness [2], and behavioral and content analyses have their own shortcomings [3].

We developed an alternative test for deception that relies on measures commonly used by cognitive psychologists to study the psychology of reading (e.g., [5]). Specifically, decisions about truth and deception were based on changes in pupil size and

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 553–562, 2009.

eye movements that occur while participants read and responded to statements about their possible involvement in mock crimes.

The size of the pupil varies with processing load and cognitive resource capacity [6]. Task-evoked pupil responses provide a reliable and sensitive psychophysiological index of the momentary processing load during performance of a wide variety of cognitive activities [7-9]. Although early research suggested that "emotional factors are relatively unimportant as determinants of pupillary responses observed in carefully controlled information-processing tasks" ([10], p. 288), an association has been noted between pupil response and emotional arousal, with larger pupil diameters associated with greater arousal [11].

Regardless of whether the effects of deception on pupil size are associated with changes in cognitive effort or emotional arousal, we predicted that participants would show greater increases in pupil diameter to statements answered deceptively than to statements answered truthfully. Consistent with our prediction, Bradley and Janisse [12] and others [13] found that pupil responses to statements on a concealed information polygraph test discriminated between truthful and deceptive subjects. Similarly, Webb and colleagues [14] found that pupil responses were as diagnostic and at least as useful as electrodermal responses for detecting deception on a comparison question polygraph test.

Results from reading studies have shown that when people experience difficulty in reading a word or phrase, their fixations on the text increase in frequency and duration, and they spend more time reading and rereading [5, 15]. Several investigators have used eye movements during reading or viewing images to detect deception. Baker, Stern, and Goldstein [16] presented test questions on a computer monitor and found that participants' fixation durations successfully discriminated between truthful and deceptive responses in 9 of 10 participants. Deceptive responses were associated with longer fixation durations than truthful responses. Other investigators were able to detect attempts to conceal information by analyzing patterns of eye movements while participants viewed images of crime locations [17] or familiar and unfamiliar stimuli [18]. Based on these findings, we predicted that deception would be associated with increased fixation frequency and reading time.

2 Method

2.1 Participants

Twenty-eight males and 12 females from the University of Utah community were recruited from fliers on campus to participate in a psychological experiment for pay and a possible bonus. Participants ranged in age from 18-38 years (M = 22.32), were predominantly Caucasian (87.5%), single (75%), and students at the University of Utah (97.5%).

2.2 Apparatus

Participants' eye movements were monitored using an Applied Sciences Laboratory (ASL) Model 501 head-mounted eye tracker. The eye tracker was interfaced with two 1.8 GHz Hewlett Packard desktop computers: one ran the eye tracker and recorded the data, and the other ran the experiment. Participants had freedom of head movement while wearing the eye tracker. Viewing was binocular, and eye movement was recorded from each participant's right eye 60 times per second. Participants' head movements and orientation were recorded with a magnetic head tracker, the output of which was stored with eye position and pupil diameter at 60 Hz.

2.3 Materials

Participants responded to 48 statements; 16 statements were neutral (e.g., "The sky is blue on sunny days."), 16 statements addressed the cash crime, and 16 statements addressed the card crime. Each statement type required an equal number of true and false responses, and each collection of true and false statements was further divided into equal numbers of statements with negation (e.g., "I did not take the $20 from the secretary's purse.") and without negation (e.g., "I took the $20 from the secretary's purse."). Eight statements were presented on eight rows on the computer monitor. The rows were 3.23 cm apart on a 54 cm monitor in portrait orientation positioned approximately 72 cm in front of the subject. Participants used a mouse to select one of two radio buttons (True or False) on the right edge of the monitor adjacent to each of the eight statements. When participants completed a page, they clicked a button at the bottom of the screen to advance to the next display that presented eight new statements. The three types of statements did not differ significantly by number of words, but they did differ by number of characters (neutral M = 38.25, cash M = 45.88, card M = 50.88; $p < .01$). To adjust for differences in statement length, number of fixations was converted to number of fixations per character, and first- and second-pass reading times were converted to ms per character.

2.4 Measures

Pupil diameter and three measures of reading were obtained for each statement and repetition. All measures were recorded only when participants had fixated within a rectangular region-of-interest that surrounded each statement. The region-of-interest was 32 mm in height, started with the first character, and ended with the last character of the statement.

Pupil diameter. Reading onset was defined as the first sample of the first of four consecutive fixations in the region-of-interest. The difference in pupil diameter between the first sample and each subsequent sample for a period of 4 s provided an evoked pupil response curve. In addition to the response curve, the area under the curve was computed to obtain a global measure of the magnitude of the pupil response. Area under the pupil response curve was the sum of positive differences between the low point that followed reading onset and each subsequent sample until the response recovered to the level of the low point or 4 s following reading onset, whichever occurred first.

Measures of reading behaviors. Number of fixations was the number of times a participant fixated in the region-of-interest. First pass duration was the time the participant spent fixating on the statement before leaving the region-of-interest and looking elsewhere. Second pass duration was the total time the participant spent rereading the statement after once having left the region-of-interest. Total time was the sum of first pass and second pass durations.

Four criteria were used to define a fixation (Eyenal Manual, Applied Sciences Laboratory, Bedford, MA). First, a fixation began at the first of six consecutive samples that occurred within .5° of visual angle. Second, any three consecutive fixation samples farther than 1° of visual angle in the horizontal or vertical direction from the running mean position ended the fixation. Third, the final fixation position was the mean position of all fixation samples between the beginning and end of the fixation period, but any two or fewer consecutive fixation samples that were farther than 1.5 standard deviations from the mean position were excluded from the calculation of the final position. Finally, any fixation duration longer than 1 second was considered an artifact and deleted.

2.5 Procedure

Participants were recruited by placing fliers for the study at various locations around the University of Utah campus. The fliers provided contact information and indicated that participants would receive pay and a potential bonus. When they called, prospective participants were given a brief description of the study, screened for inclusion criteria (i.e., over 18 years old, proficient in English, and able to read without corrective lenses), and given an appointment. Participants were then emailed initial instructions and a map of campus with a description of the study location. Participants were called the day before their scheduled appointment, reminded of their appointment, instructed to get a good night's sleep and not to drink caffeine two hours before their appointment time.

Each participant reported alone to a room on campus, entered the room, closed the door, read and signed the consent form, and read the computer-administered instructions. No researcher was present at the initial study location. After reading the instructions, the participant was given the option to discontinue the study. For those who decided to continue, they were randomly assigned to an innocent condition (n = 20) or a guilty condition. Guilty participants were further subdivided into a "Cash" crime condition (n = 10) and "Card" crime condition (n = 10). Guilty participants were informed that they had no more than 30 min to complete their assigned crime. Participants in the "Cash" crime were instructed to steal $20 from a secretary's purse. Participants in the "Card" crime were instructed to steal credit card information from a student's computer. Innocent participants were given general descriptions of the crimes but did not enact them. They were told not to report to the testing room until at least 20 minutes after the time they were scheduled to arrive for their appointment.

At the appointed time, participants reported individually to the reading lab. To motivate participants to appear truthful, they were told that in addition to a $30 payment for participating, a bonus of $30 would be paid if the participant appeared truthful to all of the statements on the test. All participants were informed that they should do their best to appear truthful on the test, and that the test was based on the idea that a person who committed a crime would have a difficult time answering. Participants' written instructions included the statement, "To appear innocent, you should respond as quickly and as accurately as you possibly can."

Participants completed a demographic questionnaire and were seated in front of the computer monitor. The ASL eye tracker was attached and calibrated. For various reasons (e.g., pupil size, eye shape, corrective lens glare), the eye movements of some participants could not be tracked (27% of prospective participants). Those individuals were paid \$15 and excused from the experiment. Participants then read a set of instructions on the computer screen. The instructions told them that statements would be presented on the computer display and they should indicate if the statement is true or false. The 48 statements were presented three times in different orders separated by an unrelated task that took 5-10 minutes to complete. The presentation of the statements was randomized across participants with the provision that a statement of one type was followed an equal number of times by a statement from the other two types and never was followed immediately by a statement of the same type.

After completion of the testing, the eye-tracker apparatus was removed. Innocent participants were paid \$60, and guilty participants were paid \$30. The participant then was debriefed about the study and asked not to share this information with anyone who might participate in the study.

3 Results

We tested two predictions: Participants would show greater increases in pupil diameter to statements answered deceptively than to statements answered truthfully, and deception would be associated with increased fixation frequency, reading time, and rereading time. Also, because statements were presented three times in different orders, we expected that deceptive participants would demonstrate less attenuation of response times across repetitions than innocent participants for statements on which they were required to lie. However, repeated-measures analysis of variance (RMANOVA) showed no meaningful effects of repetitions on any of the outcome measures.

3.1 Pupil Diameter

Pupil diameter was analyzed with RMANOVA. Time was a within-subjects factor with 40 levels (4 seconds at 10 Hz); statement type was a within-subjects factor with three levels (neutral, cash, card); and treatment condition was a between-subjects factor with three levels (innocent, cash-crime, card-crime). Huynh-Feldt corrected degrees of freedom were used for significance testing. There was a significant interaction between statement type and treatment condition, $F(4, 74) = 4.04$, $p = .005$, $\eta^2 = .18$. In addition, there was a significant three-way interaction between statement type, condition, and time, $F(76, 1406) = 2.25$, $p < .001$, $\eta^2 = .11$. Plots of the three-way interaction for each treatment condition are shown in Figure 1. Results supported the prediction that participants' pupil diameters were greatest when they read statements that they answered deceptively. Participants in the cash condition showed the greatest increase in pupil diameter in response to cash items followed by card and neutral items. Participants in the card condition showed the greatest increase in response to

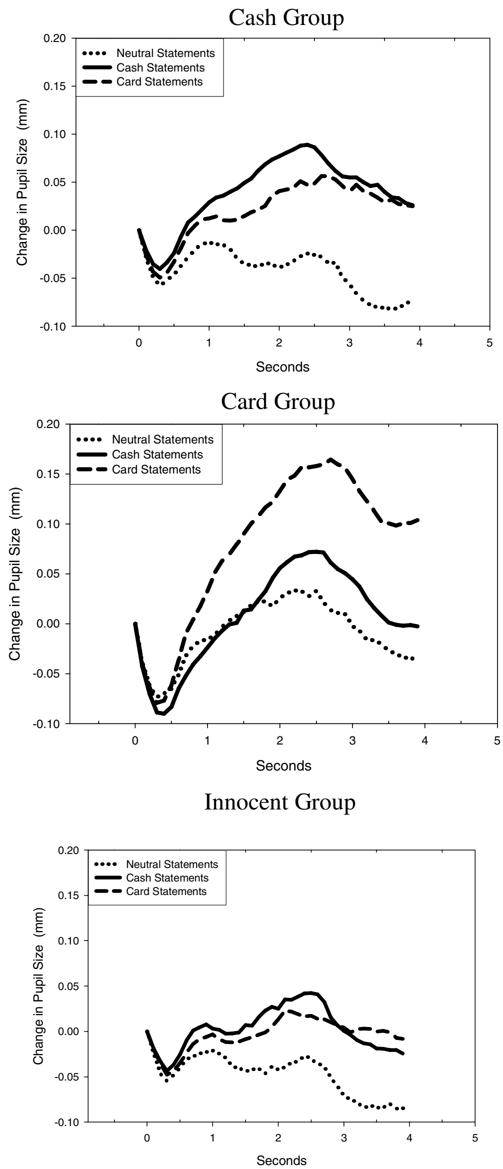


Fig. 1. Evoked pupil responses to *neutral*, *cash*, and *card statements*. *Cash statements* evoked greater increases in pupil size for participants who stole cash (*cash group*), and *card statements* evoked greater increases in pupil size for participants who took credit card information (*card group*).

card items, followed by cash and neutral items. For innocent participants, there was little difference between cash and card items, but pupil diameter was larger to those items than to the neutral items.

3.2 Reading Behaviors

There was no main effect of guilt on number on fixations per character, first pass duration (reading time), second pass duration (rereading time), or total reading time. The interaction between guilt and type of statement was significant for number of fixations per character, $F(4, 74) = 4.38$, $p = .004$, $\eta^2 = .19$; first pass duration, $F(4, 74) = 2.55$, $p = .047$, $\eta^2 = .12$, and total reading time, $F(4, 74) = 4.19$, $p = .005$, $\eta^2 = .18$, but not second pass duration. Effects on total reading time are presented in Figure 2. As expected, number of fixations and first pass duration were correlated with total reading time, and the pattern of results obtained for those three measures were similar.

Contrary to predictions, participants in the cash condition made *fewer* fixations and spent *less* time on cash than card items. Likewise, participants in the card condition made *fewer* fixations and spent *less* time on card items than cash items. Innocent participants showed little difference among cash, card, and neutral items.

3.3 Discriminant Analysis

The outcome measures included pupil diameter, response time from first fixation, number of fixations, first pass, second pass, and total time. For each of the six outcome measures, we generated three potential predictor variables. One was the participant's mean for neutral items. Another was the difference between the mean for crime-related items (cash and card) and the mean for neutral items. The last variable was the difference between the mean for cash items and the mean for card items. To find an optimal subset of variables to discriminate among innocent, cash, and card groups, two preliminary step-wise discriminant analyses were conducted. One analysis selected variables to maximize the separation between guilty and innocent participants. The second analysis selected variables to maximize the separation between the two guilty groups (cash and card). The selected variables then were included in a final multiple group discriminant analysis. This analysis yielded two discriminant functions.

Five predictor variables were included in the two discriminant functions. The first function discriminated card participants from cash and innocent participants. It relied on the difference between the cash and card statements in total reading time, the difference between the crime-related and neutral statements in pupil diameter, and the difference between cash and card statements in pupil diameter. The second function discriminated between the cash and innocent participants. It relied on the difference between cash and card statements in total reading time, the difference between the crime-related and neutral statements in rereading time, and the difference between crime-related and neutral statements in pupil diameter.

The mean classification accuracy was 78.3%. In the present study, chance accuracy was 37.5%. Ninety-five percent of innocent participants, 80% of card participants, and 60% of cash participants were classified correctly. The discriminant functions had the most difficulty discriminating between cash and innocent participants. An internal validation of the discriminant analysis (jackknife) yielded 58.3% correct classifications. Seventy-five percent of innocent participants, 70% of card participants, and 30% of cash participants were classified correctly.

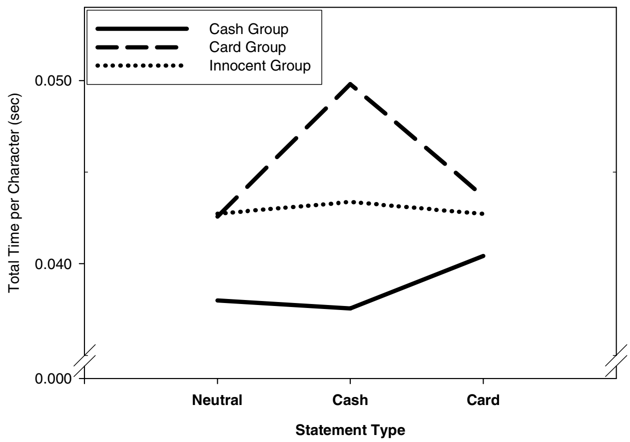


Fig. 2. Total time per character for participants guilty of taking cash (*Cash Group*), downloading credit card information (*Card Group*), or innocent of both mock crimes (*Innocent Group*). Participants in the *Cash Group* spent more time on *card* statements than cash statements. Conversely, participants in the *Card Group* spent more time on *cash* statements than *card* statements.

4 Discussion

Our purpose was to develop and evaluate a new test for detecting deception. We obtained oculomotor measures from guilty and innocent participants while they read and responded to neutral or crime-related statements on a computerized questionnaire. We predicted that participants would show greater increases in pupil diameter to statements answered deceptively than to statements answered truthfully, and deception would be associated with increased fixation frequency, reading time, and rereading time. Only the first prediction was supported. Consistent with other research [12-14], deception was associated with increased pupil size. The increase in pupil size may have been due to increased cognitive load or emotional arousal. Contrary to predictions, deception was not associated with increased fixation frequency or reading time. Rather, guilty participants made *fewer* fixations and spent *less* time reading crime-related statements answered deceptively than crime-related statements answered truthfully. These findings were obtained in the present experiment for both the cash-crime group and the card-crime group, and the findings have since been replicated in two other experiments [19-20].

Innocent participants spent about the same amount of time on all test items. In contrast, guilty participants spent significantly less time on the statements about the crime they had committed than about the crime they did not commit. The pupil data suggest that guilty participants invested more mental effort in the processing of statements answered deceptively than truthfully, and the behavioral data indicate that they were successful in reducing the time they spent reading the statement. In fact, they were too successful; they overcorrected. If overcorrection was a consequence of a conscious decision to read the incriminating items faster, it should be possible to train individuals to avoid detection by adopting a more general strategy to respond to all test items similarly.

On the other hand, overcorrection may be a consequence of a cognitive distortion known as a contrast effect or a salience effect [21]. Because the person's goal is to avoid detection when they lie, items answered deceptively stand out against a background of items answered truthfully. It has been argued that cognitive biases have evolved to facilitate information processing and protect the organism from harm (e.g., [22]). The cognitive distortion hypothesis is consistent with our findings, and it explains those of others (e.g., [23]). If it is correct, then the addition of multiple categories of items answered truthfully should increase the salience of items answered deceptively and the behavioral effects. In addition, training to defeat the test may have little effect on reading behaviors.

Oculomotor and reading measures were used to develop discriminant functions that discriminated among the three treatment conditions. The analysis yielded a mean correct classification percentage of 78.3% and a mean jackknifed classification percentage of 58.3%. The classification accuracies for the original discriminant analysis are comparable to those obtained for polygraph testing in screening contexts, but the jackknifed classification accuracies are not. The lukewarm results may have been due to the presentation of test items in a traditional questionnaire format. We presented eight statements on the computer monitor at once. Participants were free to reread statements, read statements out of order, and change their answers. Subsequent experiments in which individual statements were presented serially have produced larger effects and classification accuracies that exceed 85% [19-20]. It appears that the lack of constraint inherent in a more traditional questionnaire format introduced error in our measurements and attenuated the behavioral and physiological effects.

In conclusion, the present findings suggest that reading behaviors may be used to detect deception and may supplement or provide an alternative to the polygraph or self-report. Additional research is needed to test for effects of countermeasures, assess the effects of adding items to cover multiple issues, test if similar effects can be observed in real-world screening contexts, and determine if changes in pupil size in this context reflect changes in cognitive load or emotional arousal.

References

1. Krapohl, D.J.: The Polygraph in Personnel Screening. In: Kleiner, M. (ed.) Handbook of Polygraph Testing, pp. 217–236. Academic Press, San Diego (2002)
2. Office of Technology Assessment: The Use of Integrity Tests for Pre-employment Screening (OTA-SET-442). U. S. Government Printing Office, Washington, DC (1990)
3. Vrij, A.: Detecting Lies and Deceit: Pitfalls and Opportunities, 2nd edn. John Wiley & Sons Ltd, West Sussex (2008)
4. National Research Council: The Polygraph and Lie Detection. The National Academies Press, Washington, DC (2003)
5. Rayner, K.: Eye Movements in Reading and Information Processing: 20 Years of Research. Psychol. Bull. 124, 372–422 (1998)
6. Kahneman, D.: Attention and Effort. Prentice-Hall, Englewood Cliffs (1973)

7. Beatty, J., Wagoner, B.L.: Pupillometric Signs of Brain Activation Vary with Level of Cognitive Processing. Science 199, 1216–1218 (1978)
8. Kahneman, D., Beatty, J.: Pupil Diameter and Load on Memory. Science 154, 1583–1585 (1966)
9. Schluroff, M.: Pupil Responses to Grammatical Complexity of Sentences. Brain Lang. 17, 133–145 (1982)
10. Beatty, J.: Task-evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources. Psychol. Bull. 91, 276–292 (1982)
11. Stern, R.M., Ray, W.J., Quigley, K.S.: Psychophysiological recording, 2nd edn. Oxford University Press, New York (2001)
12. Bradley, M.T., Janisse, M.P.: Accuracy Demonstrations, Threat, and the Detection of Deception: Cardiovascular, Electrodermal, and Pupillary Measures. Psychophysiology 18, 307–315 (1981)
13. Lubow, R.E., Fein, O.: Pupillary Size in Response to a Visual Guilty Knowledge Test: New Technique for the Detection of Deception. J. Exp. Psychol. Appl. 2, 164–177 (1996)
14. Webb, A.K., Honts, C.R., Bernhardt, P., Kircher, J.C., Cook, A.E.: Effectiveness of Pupil Diameter in a Probable-lie Comparison Question Test for Deception. Legal Criminol. Psychol. (in press)
15. Rayner, K., Pollatsek, A.: The psychology of reading. Prentice-Hall, Englewood Cliffs (1989)
16. Baker, L., Stern, J.A., Goldstein, R.: The Gaze Control System and the Detection of Deception. Final report to the U.S. Government (Contract #90-F131400). Washington University, Department of Psychology, St. Louis (1992)
17. Ellson, D.G., Davis, R.C., Saltzman, I.J., Burke, C.J.: A Report of Research on Detection of Deception. Technical report prepared for Office of Naval Research (Contract N6onr-18011). Indiana University (1952)
18. Althoff, R.R., Cohen, N.J.: Eye-Movement-Based Memory Effect: A Reprocessing Effect in Face Perception. J. Exp. Psychol. Learn. 25, 997–1010 (1999)
19. Osher, D.: Multimethod Assessment of Deception: Oculomotor Movement, Pupil Size, and Response Time Measures. Dissertation Abstracts International (2007)
20. Webb, A.K.: Effects of Motivation, and Item Difficulty on Oculomotor and Behavioral Measures of Deception. Unpublished dissertation, University of Utah, Department of Educational Psychology (2008)
21. Guttel, E.: Overcorrection. Georgetown Law J. 93, 241–283 (2004)
22. Haselton, M.G., Nettle, D., Andrews, P.W.: The Evolution of Cognitive Bias. In: Buss, D.M. (ed.) Handbook of Evolutionary Psychology, pp. 724–746. Wiley, Hoboken (2005)
23. Meijer, E.H., Smulder, F.T.Y., Johnston, J.E., Merckelbach, H.L.G.J.: Combining Skin Conductance and Forced Choice in the Detection of Concealed Information. Psychophysiology 44, 814–822 (2005)

Physiological-Based Assessment of the Resilience of Training to Stressful Conditions

M. Zotov[1], J. Forsythe[2], V. Petrukovich[3], and I. Akhmedova[1]

[1] Saint Petersburg State University
{Mvzotov,pemphix}@mail.ru
[2] Sandia National Laboratories
jcforcy@sandia.gov
[3] Russian Military Medical Academy
Petrukov_vm@mail.ru

Abstract. Russian applied psychophysiology has a wide experience of using the heart rate variability (HRV) measures for the assessment of operator workload. However, 'workload indexes' that have received a wide practical application, such as tension index (TI), are not sensitive to the moment-to-moment changes of operator physiological arousal level during the performance of cognitive tasks. In this connection, a new method of HRV analysis called CS-index is offered. This index permits to identify moment-to-moment changes of operator's functional state. The presented research shows that CS-index is sensitive to task load factors, such as task difficulty level and stressful conditions and allows to differentiate experienced and novice operators during their performance on a simulator. If the CS-index proves to be reliable enough, its combination with the Automated Expert Modeling for Automated Student Evaluation (AEMASE) approach can considerably raise the efficiency of operator training.

Keywords: heart rate variability, cognitive workload, training.

1 Introduction

Russian applied psychophysiology has accumulated a rich experience of using the heart rate variability (HRV) measures for the assessment of operator cognitive and emotional workload during professional activity [2-5, 6]. Physiological hardware-software complexes such as the "Physiologist-M", used as devices to control the success of flight skills development on simulators were already used in the seventies of the XX-th century in pilot training in Russia (USSR). Such complexes were intended for pilot workload assessment ("the physiological cost of performance") at certain stages of the flight (takeoff, landing, fighting application, piloting systems failure etc.) and were based on the number of psychophysiological measures, such as heart rate, respiratory rate, respiratory minute volume, quality of secondary cognitive task performance during the performance of the basic training task [14]. Various methods of HRV analysis were developed. They enable us to assess cognitive and emotional workload level of a human operator during the professional work performance [6, 9, 12, 15].

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 563–571, 2009.

A so-called index of regulatory systems tension by Baevskii (IT) based on the analysis of characteristics of R-R intervals distribution for 3-minutes time intervals has received a wide practical application [4, 5]. The IT growth points at the increasing tone of the sympathetic branch of the autonomic nervous system, which, in turn, shows the level of mobilization of physiological regulation system on the estimated time interval. However, the IT and other similar methods of HRV analysis are not sensitive to moment-to-moment changes of operator physiological arousal level during the performance of cognitive tasks, and this complicates their usage for augmented cognition applications.

In this connection, a new method of HRV analysis called CS-index is offered [7]. This index enables to identify moment-to-moment changes of operator functional state. The approach based on the analysis of "transitive" (unsteady) phases in instantaneous heart rate (IHR) dynamics under cognitive workload intensity change was offered along with it. "Transitive" phase is understood as the period of transition of HR regulation system from one steady state to another. In the instantaneous heart rate dynamics with increased workload intensity a number of consecutive phases can be identifies: "steady-state 1" - an initial state of heart rate regulation, that is observed before the increased workload intensity; "transitive (unsteady) state"- the state of primary mobilization, characterized by IHR increasing, caused by the workload intensity increase; "steady-state 2"- a state of heart rate regulation, characterized by IHR indicators stabilization at a new level. It is indicated that in the course of training reduction of intensity and duration of "transitive" phases in IHR dynamics and decreased physiological cost of adaptation to the workloads is identified among students [11, 13].

2 Study

The purpose of the present research was to empirically assess the ability of various heart rate regulation indicators to identify changes in cognitive activity of novice and experienced operators during their professional tasks performance on simulator.

Participants. A total of 40 participants, 21 experienced train drivers and 19 novice train drivers, aged from 20 to 43 years, took part in this study. The experienced drivers were recruited from locomotive depots. The novice train drivers were recruited from the training centers of the "Russian railways" Public Corporation. The experienced train drivers had a mean train driving experience of 10 years. The novice train drivers had no experience of driving.

Procedure. A medium-fidelity fixed-base train simulator, developed by the company "Spectrum", Russia, was used in the experiment. Each participant performed six 15-minutes scenarios on a simulator. Train driver activity at the performance of the following professional tasks was simulated in the course of the scenario performance: continuous monitoring of the visual signals appearing from out of cab space, regulation of the locomotive movement speed, speaking with the use of communication device. Scenarios had three complexity levels: low, medium, high, which varied in intensity of perceived road signals and also in speed and characteristics of simulated train movement. Half of scenarios included critical incidents that simulated damages in the locomotive alarm system. The given incidents were connected with sharp increase of perceptive and working memory loads and also required operator multitask under time pressure. Simultaneously with the primary driving task participants performed a secondary sensory-motor task: they responded to visual signals appearing on the bottom of the simulator display with the frequency of 1 time in 10 seconds.

Errors made by participants during the performance of the primary driving tasks and response time for a secondary task were measured for each scenario. Time density of the sensory-motor reaction distribution was described by the formula #1 with five parameters:

$$P(T) = A \cdot (T - T^0)^B \cdot \exp\left(-C \cdot (T - T^0)^D\right). \tag{1}$$

where T^0 is an excess factor,
A - a scale factor
B –a density of distribution increase factor,
C and D – a density of distribution decrease factor.

For each examinee quantities of the mentioned factors were calculated as a result of approximation. After that the following measures were defined: (1) mode of distribution (the most probable size, TM) of reaction time values; (2) half-width of distribution (characterizes disorder of quantities, T) of reaction time quantities. Secondary sensory-motor task performance measure (SST) was calculated by the formula #2:

$$SST = 0.5 \cdot TM + 0.5 \cdot \Delta T. \tag{2}$$

The experiment has begun with the 3-minutes registration of an electrocardiogram for the baseline heart rate variability (HRV) assessment. After that participants have performed three 15-minutes scenarios, then, after a small break - three more scenarios. After the end of the last scenario 3-minute HR registration in the rest period was carried out.

Cardiovascular Measures. Electrocardiogram signals were continuously recorded during the performance of training scenarios by the participants using “Omega-M” portable loggers (Dinamica Inc., Russia) with disposable electrodes. A standard three-electrode configuration was used, as described by Mulder et al. [10]. R-peaks were continuously recorded, with an accuracy of 1 ms. Artifacts were corrected using interpolation. Analyses were made according to the recommendations of the task force on HRV [8]. Time domain measures of HRV included mean RR interval (RRNN), standard deviation of all normal RR intervals (SDNN), and RMSSD (square root of the mean squared difference of successive normal RR intervals). Frequency domain measures of HRV were quantified through the fast Fourier transform and included low frequency power (LF, 0 Hz), high frequency power (HF, 0 Hz), and the LF/HF ratio. Along with it the calculation of following measures was carried out [4]:

Mode (Mo) - most frequently occurring value of R-R.

Amplitude of a mode (AMo) – a ratio of RR-intervals quantity with the values equal to Mo, to the total RR-intervals in percentage.

Range (ΔX) - is calculated as a difference between maximum and minimum values of R-R. It reflects the variability level or peak-to-peak value of RR-intervals.

Index of autonomic balance (IAB)

$$IAB = \frac{AMo}{\Delta X} \quad (3)$$

Index of regulatory systems tension (TI) reflects a level of centralization of heart rate control

$$TI = \frac{AMo}{2 \cdot \Delta X \cdot Mo} \quad (4)$$

The analysis of the heart rate transitive processes under increased workload intensity was carried out by the technique offered by N.I. Sapova [13]. Calculation of peak and time measures of the heart rate transitive processes was carried out.

To assess moment-to-moment changes in HR-variability during the performance of training scenarios **CS-index** calculation was carried out with the formula #3:

$$CS_N(t) = \frac{\langle RRNN \rangle_A \cdot \langle SDNN \rangle_A}{\langle RRNN \rangle_N \cdot \langle SDNN \rangle_N} \quad (5)$$

Where RRNN - stands for average cardio-intervals value for an observation stage;
SDNN – standard cardio-intervals deviation for an observation stage;
<> - averaging on time interval;
t – observation time (position of the window centre);
a - baseline state (3 minutes);
N - averaging procedure for a window which width makes N points. In these calculations CSN (t) the width of a window (N) is equal to 9 points of RR-intervals.

3 Results

3.1 Task Performance Data

At the stage of data analysis we have confirmed that professional experience and a level of task difficulty and critical rail incident appearance affects the successfulness of participant performance of the primary training task and secondary sensory-motor task.

The percentage of errors made by the participants was calculated for each scenario. Measure of a primary training task's (PTT) successfulness performance was calculated on its basis. ANOVA with Task Difficulty (low, medium, high) and Conditions (presence vs. absence of critical incidents) as a within-subject factors and Experience (Novice vs. Experienced train drivers) as a between-subject factor were used. Successful performance of scenarios with critical incidents demanded from participants multitask performance during time pressure and was connected with sharp increase of cognitive workloads. In this connection we considered Conditions factor as one of the task load factors, along with the Task Difficulty factor.

It was discovered that both task load factors affect the successfulness of the primary training task performance. Increase of the training task difficulty leads to deterioration of participant activity and increase in the number of errors (F(2,37) =28,2, p<0,001). In the presence of critical incidents train drivers made more errors than in their absence (F(1,38) =16,00, p<0,001). The significant differences between the groups were determined for driving task performance during stressful scenarios, but not for neutral ones. The novice train drivers have committed more errors in driving in association of critical incidents than experienced train drivers.

Measures of the secondary sensory-motor task performance were analyzed by the same method. Again, both of the task load factors have significantly affected the performance of the sensory-motor task. The response time both for the experienced train drivers and for the novice train drivers has essentially increased in association with critical incidents, indicating the psychological cost of adaptation to stressful conditions (F(1,38)=18,9, p<0.001). Significant differences between the groups were also found (F(1,38) =14,8, p<0,01). The experienced train drivers were more successful in performing the secondary task than the novice train drivers. These results indicate that both task load factors were strong enough to significantly affect task performance. They also indicate that during the performance of training scenarios the majority of participants were sensitive to each of the task load factors.

3.2 Cardiovascular Data

Turning to the heart rate variability indicators, we have evaluated the ability of each indicator to detect changes in cognitive activity of experienced and novice drivers as each of the two task load factors varied. Each physiological measure was analyzed by the same method as the performance measures. It was found that among the measures of heart rate regulation CS measure had the greatest sensitivity to the influence of task load factors. This indicator enables to assess changes of heart rate variability for relatively small time intervals (7-9 seconds). An example of the given indicator's changes at occurrence of the critical incident is presented on the figure #1. Increase in the value of the indicator points at the increased human physiological arousal.

It was found that the Task Difficulty factor significantly affects the CS measure (F(2,37)=44,7, p<0,001). As the conditions of the training task become more complicated, the value of CS-index considerably increases in both groups of

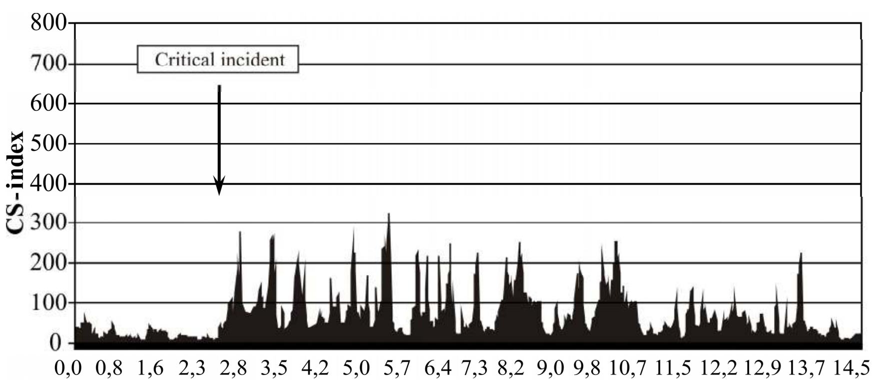


Fig. 1. An example of CS-index dynamics during the performance of stressful training task

participants. During the performance of training tasks associated with critical incidents, experienced and novice train drivers have shown greater values of CS than at the performance of neutral training tasks (F(1,38)=133, p<0,001). Significant interaction of the Task Difficulty and the factor Experience was observed (F(2,37)=7,44, p<0,05). Experienced train drivers in comparison with novice ones had considerably lower level of physiological arousal at the performance of difficult training tasks. It indicates the lower physiological 'cost' of their adaptation to raised cognitive workloads. Results of the analysis are presented at the figure #2.

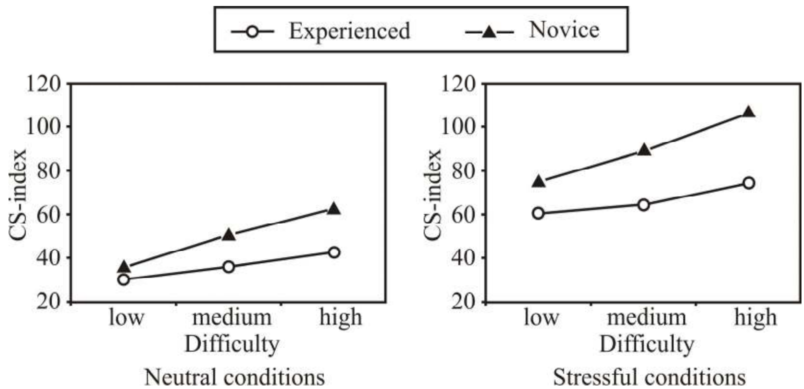


Fig. 2. Influence of Task Difficulty (low, medium, high) and Conditions (neutral, stressful) on CS measure at experienced and novice train drivers

It was established then that a number of HVR-measures offered by R.M. Baevskii [2-5] show relative sensitivity to the influence of load factors. The Condition factor (neutral, stressful) significantly affects the values of the regulatory system tension index (TI) (F(1,38)=25,06, p<0,001) and values of index of autonomic balance (IAB) (F(1,38) =33,4, p<0,001) in both groups of examinees. Measures of the regulatory system tension during the performance of neutral and stressful training tasks are presented on a figure 3. As follows from fig.3, participants had considerably greater values of a TI-index during the performance of stressful training tasks than during the performance of neutral training tasks. The influence of the Task Difficulty factor on the values of TI measures was observed (F (2,37) =5,44, p<0,05). Significant influence of the Difficulty factor on the values of IAB measures is not found (F (2,37) =2,178, p>0,05).

Significant influence of task load factors on frequency domain measures of HRV was not observed during the research.

Further, heart rate transitional characteristics during the occurrence of stressful critical incidents have been analyzed for stressful training tasks. Significant between - group differences in the measures of heart rate transitive processes duration were found (F(1,38) =9,84, p<0,01). Experienced train drivers in comparison with novice ones had considerably shorter duration of the heart rate transient phases during adaptation to high cognitive workloads caused by the occurrence of critical incidents. Influence of the Task Difficulty factor on the measures of transient processes of the heart rate was not significant.

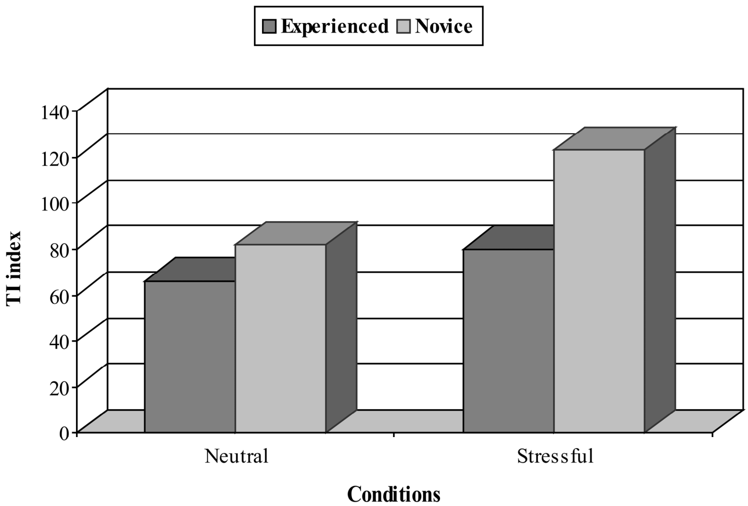


Fig. 3. Average measures of the regulatory systems tension index (TI) during the performance of neutral and stressful training tasks

4 Discussion

Russian researchers have developed a substantial number of physiological methods to assess the efficiency of operator adaptation to high cognitive workloads. Indicators of heart rate regulation (HRR) got the greatest practical application. Methods of variational pulsometry, such as regulatory system tension index (I), a measure of regulation processes adequacy (MRPA), index of autonomic balance (IAB) and other methods are widely applied in operator cognitive workloads assessment during the real or simulated professional work [2, 4]. These indexes are based on the analysis of characteristics of R-R intervals distribution. They are assessed for 3-minute time intervals and do not enable to identify moment-to-moment changes in operator cognitive activity.

In this context a new method of HRV analysis called CS-index is offered [7]. It enables to reveal moment-to-moment changes of operator's physiological arousal level during the performance of professional tasks. The approach based on the analysis of heart rate unsteady characteristics was offered along with it, enabling to assess the success of student's adaptation to specific kinds of workloads.

The purpose of the present research was to empirically assess the ability of various heart rate measures to identify changes in cognitive activity of novice and experienced drivers during the professional task performance.

It was found that during the performance of training scenarios associated with stressful incidents participants had considerably greater values of the regulatory system tension index (TI) than during the performance of scenarios not connected with stress. It confirms the data of other studies indicating that the TI index is a reliable sign of the emotional stress level experienced by operator during the work performed [5].

It was discovered, that the indicator to assess a heart rate variability called the CS-index is more sensitive to the influence of both workload factors. Increase of the training task complexity and occurrence of critical incidents led to considerable increase of the CS-index average values in both groups of participants.

The present research has also confirmed the dynamic measure informativity, such as duration of the heart rate unsteady processes during changing workloads. It was shown that experienced driver measures of unsteady processes duration were significantly less than in the novice group. It points to their more successful adaptation to the influence of the raised workloads.

It is obvious that the further research directed to validity and reliability demonstration of the considered methods of heart rate regulation assessment is necessary. If their validity and reliability are proven, it may be perspective to integrate the given methods with the Automated Expert Modeling for Automated Student Evaluation (AEMASE) approach developed by Sandia National Laboratories [1]. This will permit to raise the efficiency of expert training.

References

1. Abbott, R.G.: Automated expert modeling for automated student evaluation. Intelligent Tutoring Systems, 1–10 (2006)
2. Baevskii, R.M.: Analysis of Heart Rate Variability in Space Medicine. Human Physiology 28(2), 202–213 (2002) (in Russian); Баевский Р.М. Анализ вариабельности сердечного ритма в космической медицине. Физиология человека, 2, с. 34–43 (2001)
3. Baevskii, R.M.: Prediction of states on the verge of norm and pathology. Medicine Publ., Moscow (1979) (in Russian); Баевский Р.М. Прогнозирование состояний на грани нормы и патологии. Москва: Медицина (1979)
4. Baevskii, R.M., Ivanov, G.G.: Heart rate variability: theoretical aspects and possibilities of clinical application. Medicine Publ., Moscow (2000) (in Russian); Баевский Р.М., Иванов Г.Г. Вариабельность сердечного ритма: теоретические аспекты и возможности клинического применения. Москва: Медицина (2000)
5. Baevskii, R.M., Kirillov, O.I., Kletskin, S.Z.: The mathematical analysis of changes of a cardiac rhythm at stress. Nauka Publ., Moscow (1984) (in Russian); Баевский Р.М., Кириллов О.И., Клецкин С.З. Математический анализ изменений сердечного ритма при стрессе. М., Наука (1984)
6. Bogomolov, A.V., Gridin, L.A., Kukushkin, Y.A., Ushakov, I.B.: The diagnostics of human state: Mathematical Approaches. Medicine Publ., Moscow (2003) (in Russian) Богомолов А.В., Гридин Л.А., Кукушкин Ю.А., Ушаков И.Б. Диагностика состояния человека: математические подходы. - М.: Медицина (2003)
7. Bozhokin, S.V., Petrukovich, V.M., Negrusha, N.A., Inozemtzev, S.A.: Vegetative modulation of heart rate under psychophysiological loads. Vestnik Rossiiskoi voenno-meditsinskoi akademii (News of Russian Military-Medical Academy) 1(17), 332 (2007) (in Russian); Божокин, С.В., Петрукович, В.М., Негруша, Н.А., Иноземцев, С.А. Вегетативная модуляция сердечного ритма при психофизиологической нагрузке // Вестник Российской Военно-медицинской академии, приложение 1(17), С. 332 (2007)
8. Camm, A.J., Malik, M., Bigger, J.T., Breithardt, G., Cerutti, S., Cohen, R.J., Coumel, P., Fallen, E.L., Kennedy, H.L., Kleiger, R.E., Lombardi, F., Malliani, A., Moss, A.J., Rottman, J.N., Schmidt, G., Schwartz, P.J., Singer, D.H.: Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Eur. Heart J. 17, 354–381 (1996)

9. Medvedev, V.I.: Human's physiological and psychological functions' resistance under the influence of extreme factors. Science Publ., Leningrad (1982) (in Russian); Медведев, В.И.: Устойчивость физиологических и психологических функций человека при действии экстремальных факторов. Ленинград, Наука (1982)
10. Mulder, L.J.M., de Waard, D., Brookhuis, K.A.: Estimating mental effort using heart rate and heart rate variability. In: Stanton, N., Hedge, A., Brookhuis, K.A., Salas, E., Hendrick, H. (eds.) Handbook of Human Factors and Ergonomics Methods. CRC Press, Boca Raton (2004)
11. Novikov, V.S., Shustov, E.B., Goranchuk, V.V.: Correction of functional states under extreme exposures. Science Publ., St. Petersburg (1998) (in Russian); Новиков, В.С., Шустов, Е.Б., Горанчук, В.В. Коррекция функциональных состояний при экстремальных воздействиях. - СПб.: Наука (1998)
12. Parin, V.V., Baevskii, R.M.: Methods of heart rate's mathematical analysis. Science Publ., Moscow (1968) (in Russian); Парин, В.В., Баевский, Р.М. Математические методы анализа сердечного ритма. Москва: Наука (1968)
13. Sapova, N.I.: Complex evaluation of heart rhythm regulation during measured functional loads. Fiziologicheskiĭ zhurnal SSSR imeni I. M. Sechenova (Sechenov physiological journal of the USSR) 68(8), 1159–1164 (1982) (in Russian); Сапова, Н.И. Комплексная оценка регуляции сердечного ритма при дозированных функциональных нагрузках // Физиол. журн. им. Сеченова 68(8), С. 1159–1164 (1982)
14. Zhernavkov, V.F., Kozlovsky, E.A.: Psychophysiological assessment of pilots' readiness on the flight simulators. Voenizdat (Military Publ. House), Moscow (1981) (in Russian); Жернавков, В.Ф., Козловский, Э.А. Психофизиологическая оценка подготовленности летчика на пилотажных тренажерах. Москва: Воениздат (1981)
15. Zotov, M.V., Petrukovich, V.M., Zhuravleva, O.P., Inozemtzev, S.A., Voit, A.P.: The automated systems of dynamic control of the operators' emotional state and professional working capacity. Vestnik Rossiiskoi voenno-meditsinskoi akademii (News of Russian Military-Medical Academy). 1 (15), pp. 442-443 (2006) (in Russian) Зотов, М.В., Петрукович, В.М., Журавлева, О.П., Иноземцев, С.А., Войт, А.П. Автоматизированная система динамического контроля функционального состояния и профессиональной работоспособности операторов. Вестник Российской Военно-медицинской академии, приложение, 1(15), С. 442–443 (2006)

Tunnel Operator Training with a Conversational Agent-Assistant

Eric Buiël, Jan Lubbers, Willem van Doesburg, and Tijmen Muller

TNO Defense, Security and Safety, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands
{eric.buiel,jan.lubbers,willem.vandoesburg,tijmen.muller}@tno.nl

Abstract. A tunnel operator monitors and regulates the flow of traffic inside a tunnel. Tunnel operators need to train in a simulator regularly in order to maintain proficiency in handling incident situations. During quiet working hours, the operator has enough time for training. But generally at that time no instructor or colleague operators are present to provide instruction, advises, and feedback. As a solution, we have designed an automated training system. The system employs a conversational agent which supports the operator's situation assessment tasks. The agent exhibits peer behavior which is unobtrusively directed by didactic strategies. In this paper we present the design, development and application of the agent.

Keywords: Agent-Based Modeling and Training, Cognitive Modeling, Constructive Learning, Intelligent Virtual Agent.

1 Introduction

The task of a tunnel operator is to monitor and regulate the flow of traffic inside a tunnel. Surveillance cameras are mounted inside the tunnel. Images from the cameras are displayed on monitors inside the control room. In case of a large-scale incident, the operator takes initial safety measures (e.g. closing traffic lanes) and stays in charge; at least until the principle fire-department officer arrives on the scene.

Tunnel operators need to train regularly in order to maintain proficiency in developing situation awareness and handling the events occurring in the stressful initial minutes of a crisis situation. We have developed a training simulator that enables a tunnel operator to train himself in incident management. The simulator presents an incident-situation to the operator, which develops in real-time. The operator has to make situation assessments and must take appropriate safety measures.

Since incidents occur infrequently, the tunnel operator's workload is unevenly distributed over time. During quiet working hours, the operator has enough time for simulator training. But generally at that time no instructor is present to provide didactic support (e.g. providing instructions and feedback). No other operators are present to provide colleague advises either. Therefore, educational support systems need to assist the operator during the training. Intelligent virtual agent technology enables the creation of virtual characters that conduct a true dialogue with humans [1] [2]. Intelligent virtual agent technology combines a realistic real-time three-dimensional

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 575–584, 2009.

visualization of a human, with the cognitive modeling of 'what is on the human's mind', and means to communicate in spoken and written word. The application of agent technology enables us to create virtual agents that provide didactic support to the tunnel operator, for example virtual instructors and virtual co-learners. Yet, not much is known about the most effective way of applying these virtual agents for operator training.

2 Research Goals and Approach

In this paper we present the design, development and application of an educational agent for the training of a tunnel operator. The main research question of this project was: What are the necessary agent functionalities for virtual agents that accompany a tunnel operator during task training in a tunnel training simulator? Here, we considered knowledge requirements (i.e. which domain knowledge, didactic knowledge, and student knowledge has to be available for the agent?) as well as behavioral requirements (i.e. what is the most effective content, form and timing of the agent's interventions and instructions?). We first performed a literature study on agent-based training applications. Based on the results of this study, we composed a generic framework that describes the cooperation between a human supervisor and an educational agent jointly training a supervisory task, like the tunnel operator task. Based on this framework, we implemented a prototype educational agent in our tunnel training simulator. In future projects, this prototype can be used to refine the agent design guidelines by means of training experiments.

This paper provides an overview of the project results. §3 presents an overview of generic design guidelines for educational agents obtained from literature. §4 through §7 describe the agent-student cooperation framework and the prototype educational agent. §8 concludes with a summary and the focus of future training experiments.

3 Educational Background

In 2000, Johnson [3] recognized that animated pedagogical agents were in the early stages of development, but that they would have a significant impact on education and training in the near future [4] [5]. Engaging, expressive pedagogical agents can provide feedback and advice that have a strong motivating effect on trainees, and may even encourage and empathize with these trainees. Moreno et al. [6] conducted a study investigating the effectiveness of an animated pedagogical agent on children's learning. One group of trainees learned with on-screen instruction and the help of a pedagogical agent and another group learned just by reading on-screen instructions. Across two separate experiments, the agent group had a 24-48 % higher success rate than the non-agent group. For children, subtle praising for effort proves to provide good result; better than praising for accomplishment [7].

Another effect that can be expected from a life-like agent is that the trainee will build up a relationship with an agent and might become committed to respond to the agent in a positive way. According to Kidd and Breazeal [8], the most important factors for trying to create and maintain a helpful, long-term human-agent relationship are engagement, trust and motivation. The agent used relational strategies such as social dialogue, empathy dialogue, meta-relational communication, humor, continuity behaviors, and forms of address and politeness strategies [9]. Personality is fundamental to social relationships. People automatically perceive a personality in social agents even when no personality is intended [10]. This has led to the "*computer as a social actor*" paradigm put forward by Reeves and Nass [11]. In this paradigm all computer mediated interaction should take human social assumptions into account. Therefore, explicitly adding a personality can direct the social interaction into the desired direction.

The classical student-teacher learning method implies a hierarchical relationship between student and instructor instead of a social relationship. The instructor chooses the exercises, directs the learner, evaluates task performance, and exactly tells the learner how to act in order to improve performance. This classical method does not match sufficiently with the "computer as a social actor" paradigm. Modern, self-directed, constructive learning methods [12] [13] provide a better match. Here the trainee is fully responsible for his own task proficiency. The trainee is encouraged to improve his proficiency, preferably by experimenting and by cooperating with colleague trainees. Trainees discuss their task strategies together, and think about alternative strategies that might improve task performance. As a consequence, social interaction between trainees is crucial to the success of these learning methods.

In order to implement the above-mentioned social, constructive learning approach in an agent-based learning environment, like the tunnel training simulator, we need an educational agent acting like a co-learner. This *companion agent* encourages the trainee to discuss his thoughts and actions and stimulates discoveries on how to reach the learning goals. The agent has the same expertise as the operator does. This means that the agent can provide good suggestions, but might also make mistakes. The trainee does not have to follow the agent's advice. The agent just needs to make him think over certain situations that occur during the scenario execution.

In our tunnel training simulator, we implemented a prototype companion agent (§4) that supports the trainee in situation assessment tasks (§5). The design framework of the agent is described in §6. §7 exemplifies typical agent interventions.

4 Demonstration Setup

Figure 1 shows the tunnel training simulator with companion agent. The tunnel training simulator implements a virtual model of the tunnel control room. Instead of operating on a real tunnel control system, commands are sent to a virtual tunnel simulator and sensor signals (surveillance camera pictures, alarm messaging from automated surveillance systems) are received in return. The trainee and companion agent interact socially while executing their joint task. In Figure 1, the companion agent is indicated by ①, the virtual tunnel control system is indicated by ②, and the virtual tunnel monitor is indicated by ③. In this setup, the operator can switch easily between operating the control system, monitoring the tunnel, and interacting with his co-learner.

The companion agent can interact with the trainee through verbal communication, with either written or spoken dialogue. The former is implemented by a text based input module, the latter by a voice recognition module. The agent is embodied

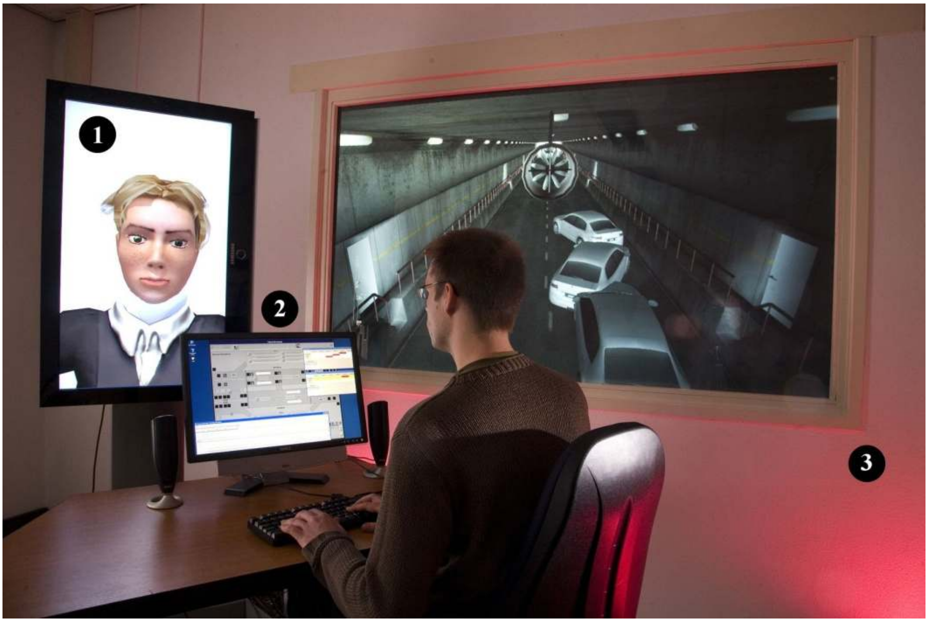


Fig. 1. Tunnel training simulator with a companion agent

through an animated 3D character model, and a speech synthesis module transforms written dialogue to speech, which is fed into a lip-synchronization module. The behavior of the agent is discussed in §6.

5 The Task of the Tunnel Operator

Tunnel incidents are categorized and ranked by their impact on tunnel safety. For each incident type, a different incident procedure exists. The tunnel operator needs to execute this procedure in order to return to a safe situation. E.g. for a simple car crash without injuries and fire, he needs to close the traffic lane and call in the tow service. For a large-scale fire incident, he needs to push an alarm button that activates automatic safety measures (e.g. activating route signing of emergency exits). Table 1 shows the incident types of the tunnel training simulator.

Table 1. Incident classifications in the tunnel training simulator

Incident type	*Incident name*	*Events occurring inside the tunnel*		
		Damage	*Injuries*	*Fire*
0	No incident	No	No	No
1	Incident with material damage	Yes	No	No
2	Incident with injuries		Yes	No
3	Fire incident			Yes

The task of the tunnel operator is largely a situation assessment task. In order to perform well, the tunnel operator must have good situation awareness (SA). Endsley [14] defines SA as: “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future”. Translated to the specific tunnel operator task, the operator needs to do *observations* (phase 1 of SA, perceptions) and use these to classify the current situation in the form of a *diagnosis* (phase 2 of SA, comprehension). Based on the diagnosis, the operator decides on a course of action (the third phase of SA, projection).

Learning objectives. From a situation awareness perspective, the learning goals of tunnel operator training are threefold:

1. The trainee is able to find the information that is necessary to make a situation assessment (the *observation objective*, related to the first phase of SA);
2. The trainee is able to classify an incident situation by combining individual observations into a correct situation assessment (the *diagnosis objective*, related to the second phase of SA);
3. The trainee has developed “critical thinking” skills [15]. From time to time, he needs to reassess the situation, and decide whether his current situation classification is still correct (the *critical thinking objective*).

In the current simulator, the learner does not have to decide on the course of action; in other words, projection is not a learning objective in this implementation.

Error types. The trainee may fail to perform correct diagnosis formation due to a number of reasons. We categorize five error types in total, divided over the learning objectives mentioned above. For the observation objective these errors are:

1. the *unseen error*: the trainee has failed to notice something;
2. the *hallucination error*: the trainee reports something that does not take place.

For the diagnosis objective, there are two typical errors:

3. the *classification error*: the trainee has not applied the incident classification rules correctly;
4. the *omission error*: the trainee has not shared an observation with his peer.

Finally, there is one typical error within the context of the critical thinking objective:

5. the *critical attitude error*: the trainee does not reassess the situation frequently, or holds on to an obsolete diagnosis too long.

6 The Cooperation between Learner and Agent

For effective constructive learning, it is vital that the trainee is able to speak freely with his co-learner, without feeling judged. This implies the absence of an authoritarian relationship. Therefore, the agent must act as a true companion to the trainee. As such, the agent is required to exhibit peer behavior to create a safe and trustworthy atmosphere. At the same time, the agent should stimulate conversation with the trainee in order to encourage him to explain his reasoning and to enable him to do new discoveries and construct new knowledge. In order to achieve this, the agent is required to guide the didactic process. Didactic guiding must occur only unobtrusively and must remain unnoticed by the trainee.

Figure 2 illustrates how these two principles (peer behavior and didactic guiding) are incorporated in the design of the agent. The agent is divided in two separate layers: the didactic layer and the peer behavior layer. The top layer (didactic module) contains domain-independent didactic knowledge. In the bottom layer (peer behavior module), the behavior of the agent is modeled to represent the role of a companion to the trainee. On this layer the specific domain knowledge is modeled. This design enables easy reuse of the agent for training other supervisory tasks. Only the specific domain knowledge in the peer behavior module needs to be remodeled.

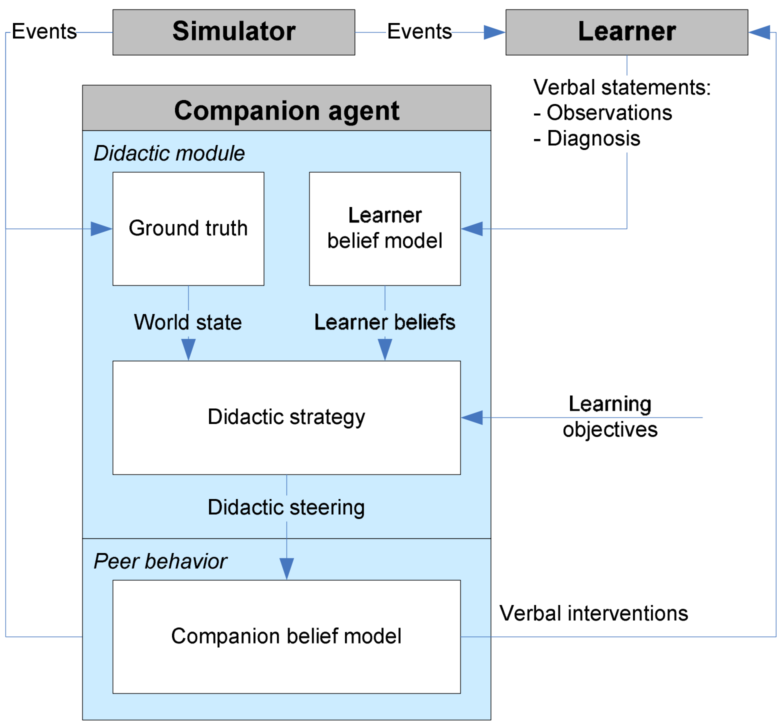


Fig. 2. Companion agent design framework

The agent's peer behavior is executed by the peer behavior module and directed by the didactic module. The didactic module has a complete view on the world, whereas the peer behavior module and the trainee only have partial (and possibly incorrect) views on the world. These views are based upon events sent by the simulator.

The trainee can share his beliefs with the agent verbally by expressing statements. Each statement can either regard a new observation (e.g. *"I see car damage on camera 5."*) or a situation diagnosis (e.g. *"There is a fire incident."*). These statements are semantically processed into beliefs, which are added to the learner belief model. The didactic module compares the learner's beliefs to the ground truth in order to recognize and classify possible errors (see §5). The result of this classification is used to direct the peer behavior of the agent. The didactic module determines if the trainee's observations are correct, sufficient and necessary to support his diagnosis. Depending on the classification, different didactic strategies can be applied. The didactic module employs several dialogue models to encourage the trainee to explain his diagnosis, including asking the trainee for supportive evidence, encouraging the trainee to state his observations, and intervening (e.g. deny an observation, suggest a review of the observation, or propose a different action).

7 Typical Agent Interventions

For each learning objective, this section describes the most typical agent interventions.

The observation objective. If the trainee's latest verbal statement indicates a new observation, the agent will respond following the flowchart in Figure 3. If the observation is incorrect, the provided verbal intervention depends on the type of observation error. For example, when the trainee incorrectly observes that injuries are present (hallucination error), the agent will focus on more detailed observation, and respond *"I'll go check the number of injuries."* Another example is when the trainee incorrectly observes that injuries are not present (unseen error), the agent will focus on typical incident characteristics, and respond *"There is a chance that people are hurt."* For each error type, multiple response phrases are available, such that the agent can select a different response phrase when the trainee holds on to an incorrect observation.

If the observation is correct, the agent will stimulate the trainee to investigate if this observation changes the diagnosis.

The diagnosis objective. If the trainee's latest verbal statement indicates a new diagnosis, the agent will first ask the trainee to explain his diagnosis by calling out the observations that support the diagnosis one at a time. Subsequently, the agent will respond following the flowchart in Figure 4.

If (some of) the supporting observations are incorrect, or if the trainee did not mention crucial information in relation to the provided diagnosis, the agent will stimulate the trainee to reconsider his observations. The agent categorizes the incorrect or missing observations as hallucination errors, unseen cue errors, and omission errors, and selects a single error from the observation set. If the selected error is a hallucination error or an unseen error, the agent uses verbal responses similar to Figure 3, in order to encourage the trainee to reconsider his diagnosis. If the selected error is an omission error the agent tells the trainee his own beliefs regarding the specific subject. For example, suppose the trainee correctly diagnosed the incident situation as an incident with injuries (incident type 2 in Table 1). When the trainee did not mention the absence of fire in his diagnosis explanation, the agent responds "*I think there is a fire.*" This statement is incorrect, and probably the trainee will correct the agent. But at the

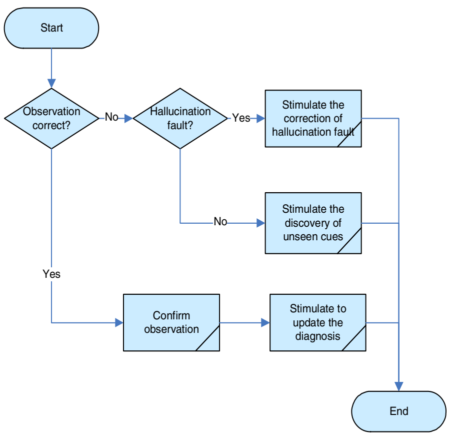


Fig. 3. Handling new observations

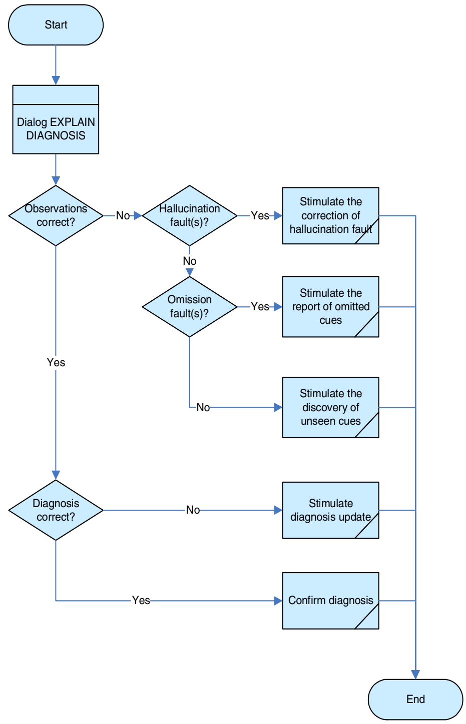


Fig. 4. Handling new diagnoses

same time, it will probably make him think over the relation between the presence/absence of fire and the diagnosis of a type 2 incident.

If the trainee's observation set is correct and complete, but his diagnosis is incorrect (classification error), the agent will stimulate the operator to reconsider his classification by providing his own diagnosis belief. Once more, this belief may even be incorrect.

The critical thinking objective. If the trainee's diagnosis belief does not change quickly after an event in the simulated tunnel (i.e. a change of the ground truth), the agent concludes the trainee holds on to an obsolete diagnosis. Then, the agent provides unobtrusive hints in order to stimulate the trainee to suspend his current activities and start a critical review of his current diagnosis belief. This is likely to occur when shortly after a first incident, a second incident occurs. If the second incident remains unnoticed by the trainee, the agent will first ask *"What is going on?"* If the trainee holds on to his initial diagnosis the agent will ask *"Which procedure did you follow?"* Subsequently he will ask *"When did it happen?"* The last-mentioned question causes the trainee to place the earlier observed events in a time frame, and encourages the trainee to investigate whether current events are in line with the events observed earlier.

8 Conclusions and Future Work

In order to apply virtual agents for training, the social relationship between the agent and the trainee has to be taken into account. The agent must consciously balance between motivating and guiding the learner. On the one hand, the agent must prominently act as a true companion to the trainee, and exhibit peer behavior to create a safe and trustworthy atmosphere. On the other hand, the agent must unobtrusively guide the didactic process: the agent must encourage the trainee to verbally explain his reasoning in order to enable him to do new discoveries and construct new knowledge. Didactic guiding must remain unnoticed by the trainee.

Our agent-learner cooperation framework (§6) enables the creation of social educational agents supporting a human supervisor when training a supervisory task. The implemented companion agent prototype for tunnel operator training shows the capabilities at hand within this framework. Future experiments have to show if training with this agent indeed results in improved situation awareness and solid incident assessments. These experiments also have to provide the most effective content, form and timing of the agent's interventions and instructions.

Acknowledgments

The research reported here is part of the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, grant nr. BSIK03024.

References

1. Rickel, J., Marcella, S., Gratch, J., Hill, R., Traum, D.R., Swartout, W.R.: Toward a New Generation of Virtual Humans for Interactive Experiences. IEEE Intelligent Systems 17(4), 32–38 (2002)
2. Traum, D.R., Rickel, J.: Embodied agents for multi-party dialogue in immersive virtual worlds. In: Proceedings of the first International Joint conference on Autonomous Agents and Multiagent systems, pp. 766–773 (2002)
3. Johnson, W.L., Rickel, J.W., Lester, J.C.: Animated pedagogical agents: Face-to-face interaction in interactive learning environments. International Journal of Artificial Intelligence in Education 11, 47–78 (2000)
4. Atkinson, R.K.: Optimizing learning from examples using animated pedagogical agents. Journal of Educational Psychology 94, 416–427 (2002)
5. Clark, R.C., Mayer, R.E.: e-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning. John Wiley & Sons. Inc., San Francisco (2003)
6. Moreno, R., Mayer, R.E., Spires, H., Lester, J.: The case for social agency in computer-based teaching: Do trainees learn more deeply when they interact with animated pedagogical agents? Cognition and Instruction 19, 177–214 (2001)
7. Mueller, C.M., Dweck, C.S.: Praise for intelligence can undermine children's motivation and performance. Journal of personality and social psychology 75, 33–52 (1998)

8. Kidd, C.D., Breazeal, C.: Sociable robot systems for real-world problems. In: IEEE International Workshop on Robot and Human Interactive Communication, pp. 353–358 (2005)
9. Bickmore, T.W., Picard, R.W.: Establishing and maintaining long-term human-computer relationships. ACM Transactions on Computer-Human Interaction (TOCHI) 12, 293–327 (2005)
10. Dryer, D.C.: Getting personal with computers: How to design personalities for agents. Applied Artificial Intelligence 13, 273–295 (1999)
11. Reeves, B., Nass, C.: The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press, Cambridge (1996)
12. van der Hulst, A., Muller, T.J., Besselink, S., Coetsier, D., Roos, C.L.: Bloody Serious Gaming – Experiences with Job Oriented Training. In: Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL (2008)
13. Stehouwer, M., Stricker, J., van Gemeren, W.: Training Design for Professional Development. In: Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL, p. 12 (2006)
14. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Human Factors 37(1), 32–64 (2005)
15. van den Bosch, K., Helsdingen, A.S.: Critical Thinking in Tactical Decision Games Training. In: Cook, M., Noyes, J., Masakowski, Y. (eds.) Decision Making in Complex Environments, pp. 213–222. Ashgate, Aldershot (2007)

Evaluating Training with Cognitive State Sensing Technology

Patrick L. Craven[1], Patrice D. Tremoulet[1], Joyce H. Barton[1], Steven J. Tourville[2], and Yaela Dahan-Marks[2]

[1] Lockheed Martin Advanced Technology Laboratories, 3 Executive Campus, Suite 600, Cherry Hill, NJ 08002
[2] Lockheed Martin Simulation, Training & Support, Orlando, FL
patrick.craven@lmco.com, polly.d.tremoulet@lmco.com, jbarton@atl.lmco.com, steven.tourville@lmco.com, yaela.dahan-marks@lmco.com

Abstract. Five different training techniques (classroom, video, game-based, computer-based, and simulator) were compared using neurophysiological measurements. The best performance was displayed by individuals in the classroom and video conditions. These participants also displayed the lowest levels of cognitive workload and the highest levels of engagement. The poorest performance on the training was exhibited by individuals in the computer-based and game conditions. These participants also displayed the highest levels of cognitive workload, the lowest levels of engagement, and computer-based had the highest levels of drowsiness. As expected, the testing phases of the training had the highest levels of workload. In general, engagement dropped and distraction increased during the training phase when the material was first presented to participants. However, participants who could keep engagement high during this period performed better. This suggests that mental state monitoring during training could help provide a mechanism for alleviating distraction and inattention and boost training efficacy.

1 Introduction

The use of neurophysiological sensors such as electroencephalograph (EEG) to measure cognitive workload is not a novel concept. Other researchers have investigated the relation between mental effort and Gamma band response [1] and power and coherence measures as they relate to mental effort [2]. The research team used a gauge that is based on aspects of these underlying responses and provides a quantified measure in real time of an individual's mental state.

The Sensor-based Mental Assessment in Real Time (SMART) system utilizes user-independent algorithms to convert physiological data into multiple gauge readings. These gauges indicate second-by-second variation in mental activity and provide insight into the participant's cognitive processing. Electroencephalograph (EEG) sensors are used to provides measurement of voltage fluctuation across the scalp, which is indicative of cognitive processing. SMART allows real-time viewing of the

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 585–594, 2009.

calculated mental state gauge values as an individual performs a task. The gauges represent a probability of being classified as a particular model. The mental state gauges include cognitive workload, distraction, engagement, and drowsiness. The former was validated in research conducted by Tremoulet et al. [4], and the other gauges have been verified through other research [5]. The cognitive workload value is derived from a two-state classifier, and the gauge output reflects the probability of the high workload state. The distraction, engagement, and drowsiness values are derived from a four-state classifier.

The SMART system has been used in a military command and control environment in which an operator's mental state classification provided feedback to an interface, which allowed the interface to adapt to changes in mental state in real-time [6]. Additionally, the SMART system was validated as a tool for improving interface design when it was validated against other measures of cognitive workload [4]. Naturally, one can imagine using this valuable mental state information in a training domain as well where trainers are keenly interested in the efficacy of various training methods as well as maintaining optimal levels of attention and minimal distraction during training scenarios.

The relationship between cognitive workload and training is ambiguous. Intuition suggests that one's workload should decrease on a task as one becomes more familiar and better trained. Novice drivers must focus a considerable amount of cognitive resources towards the task of driving, but more experienced drivers have developed automatic responses that enable them to have more available cognitive resources [7]. A study by Fournier, Wilson, and Swan investigated the changes in workload as measured by the NASA TLX, and physiology, as measured by alpha wave attenuation and theta wave enhancement [8]. Alpha waves are brain waves of approximately 8 to 12 Hz, and theta waves are approximately 4 to 7 Hz. The study revealed that alpha and theta fluctuations did not relate to the amount of training that participants received. This lack of relation may possibly be explained by the fact that participants only participated in six sessions at approximately two minutes long with the middle four sessions serving as training. Taken together, these results suggest that changes in EEG as a result of training might be observable, but currently research hasn't linked training to measurable changes in EEG waveforms.

The purpose of the current study was to determine whether neurophysiologically-based gauges could detect changes in mental state based on the effectiveness of training. Additionally, the current study would allow for the investigation of the type of changes in mental state which that throughout the course of a training episode.

2 Method

Eighteen individuals (15 males and three females) from a military service academy participated in the study. The mean age of the participants was 19.6 years. Of the participants, 12 were right-handed, four were left-handed, and two were ambidextrous.

The experiment was a between-subjects design with each subject randomly assigned to one of five training conditions. Participants were trained to identify three friendly and three enemy (threat) vehicles within one of five training conditions. The training conditions included a) game-based, b) simulator, c) computer-based training (CBT), d) video, and e) classroom. The game-based condition embedded the training within a computer game. The simulator condition embedded the training within a simulator environment. The CBT condition used flash animation to train participants. The classroom and video conditions both used an instructor; in the classroom condition the instructor was physically present while in the video condition the instructor had been recorded at an earlier time. The instructor followed a designated protocol when conducting the live classroom training.

The SMART tool was used to measure the cognitive state of participants during the study. The SMART system uses task-independent and user-independent algorithms to convert physiological data into multiple gauge readings that indicate second-by-second variation in mental activity. These mental state gauges, which include cognitive workload, distraction, engagement, and drowsiness, provided insight into the participant's mental processing. SMART provides these objective measures of mental state once per second without interrupting performance.

For this study, two functionally equivalent testing stations were established that each used three sensor types. The first two sensor types were collected using an EEG sensor cap from Advanced Brain Monitoring (ABM) that is also equipped with two EKG leads. EEG data was acquired from a wireless sensor headset developed by ABM from five channels using the following bi-polar montage: C3-C4, Cz-PO, F3-Cz, Fz-C3, Fz-PO. Bi-polar differential recordings were selected to reduce the potential for movement artifacts that can be problematic for applications that require ambulatory conditions in operational environments.

Each participant's test session lasted approximately one and one half hours. The participants received a briefing on the purpose and procedures for the study, and participant numbers were assigned to ensure participant anonymity. The participant's name was kept separate from all quantitative data.

During the five phases (pre-test, training, assessment with feedback, assessment without feedback, and post-test), the experimenter marked experiment phases. For the purpose of grouping the gauge data appropriately, the training was divided into five phases that were marked by the experimenter. A pre-test preceded Phase 2, and a post-test followed Phase 4. Phase 2 was the introduction of the six vehicle types. Phase 3 was a knowledge assessment including detailed feedback, and Phase 4 was a knowledge assessment with accuracy feedback without the additional details contained in Phase 3.

3 Results

The training lasted approximately 30 minutes per participant, of which the first 18 minutes were devoted to learning about the six vehicles (Phase 2). Phase 3 and Phase 4 were each approximately five minutes long. In order to condense the large volume of data, the average values of these second-long measurements for each task phase (including pre- and post-tests) were calculated. These four different gauge values means are summarized in Figure 1.

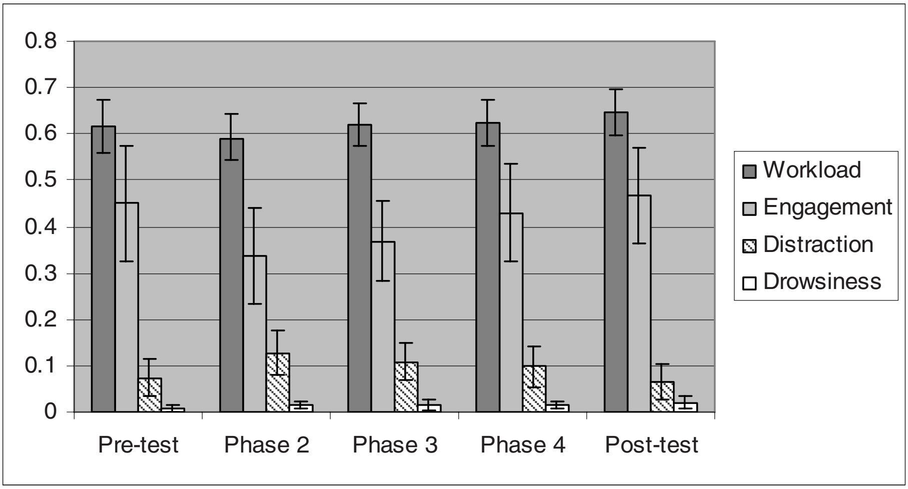


Fig. 1. Gauge value means with 95% confidence intervals

3.1 Gauge Values Variation by Task Phase

One of the principal research goals of this study was to understand how the mental state gauges varied by task phase. The following sections provide a description of the repeated measures (within subjects) analysis of variance (ANOVA) conducted for each of the four gauge means for each of the five task phases.

The repeated measure ANOVA for cognitive workload was significant ($F(4,64) = 5.40$, $p < .01$). Differences between adjacent phases were inspected using ANOVA contrasts, and the difference between Phase 2 and Phase 3 ($F(1,16) = 5.41$, $p < .05$) and the difference between Phase 4 and Phase 5 ($F(1,16) = 6.49$, $p < .05$) were significant. Although the difference between Phase 1 and Phase 2 is large, this difference failed to be statistically significant. The trend for workload was to start modestly, drop lower for Phase 2, increase towards pre-test levels in Phases 3 and 4, and ultimately climbing to its peak in the post-test.

The repeated measure ANOVA for engagement was significant ($F(4,64) = 3.28$, $p < .05$). Differences between adjacent phases were inspected using ANOVA contrasts, and no phases were statistically different. However, the difference between Phase 1 and 2 ($F(1,16) = 4.01$, $p = .06$) and the difference between Phase 3 and 4 ($F(1,16) = 4.06$, $p = .06$) were marginally significant. Although the difference between Phase 2 and Phase 3 appears to be large, this difference was not significant. The general trend for engagement was that the level started fairly high in Phase 1, dropped lower for Phase 2, and then proceeded to recover in Phases 3 through 5.

The repeated measure ANOVA for distraction was significant ($F(4,64) = 5.07$, $p < .01$). Differences between adjacent phases were inspected using ANOVA contrasts, and the difference between Phase 1 and Phase 2 was significant ($F(1,16) = 13.05$, $p < .01$). Additionally, the difference between Phase 4 and 5 ($F(1,16) = 3.87$, $p = .07$) was marginally significant. Although the difference between Phase 2 and Phase 3 appears to be large, this difference failed to be statistically significant. The general trend for distraction was that the gauge increases significantly such that Phase 2 is a higher level of distraction. Distraction then drops off from Phase 2 to the post-test. The lowest mean levels of distraction were in the pre and post-test.

The repeated measure ANOVA for drowsiness was not significant. Differences between adjacent phases were inspected using ANOVA contrasts, and the difference between Phase 1 and Phase 2 was significant ($F(1,16) = 4.55$, $p < .05$). Although it is difficult to interpret this gauge because of the lack of a main effect, drowsiness appears to increase between Phase 1 and 2, remain relatively stable from Phase 2 through 4, and then increase modestly from Phase 4 through Phase 5.

3.2 Gauge Values by Training Condition and Phase

The difference in the workload values by condition was also examined. An ANOVA was conducted for each combination of gauge value and task phase to determine whether, for example, cognitive workload differed in the pre-test phase between the game and simulation condition. The means for cognitive workload by task phase and condition are presented in Table 1. There were significant differences in cognitive workload in the pre-test ($F = 4.82$, $p < .05$), Phase 3 ($F = 4.20$, $p < .05$), Phase 4 ($F = 4.94$, $p < .05$), and post-test ($F = 3.32$, $p < .05$) phases.

Table 1. Cognitive workload means by phase and condition

	Pre-test	Phase 2	Phase 3	Phase 4	Post-test
Game	.713	.670	.694	.709	.718
Simulator	.590	.589	.615	.619	.638
CBT	.741	.669	.700	.708	.729
Video	.515	.478	.512	.523	.514
Classroom	.548	.547	.559	.549	.598

Fisher's least significant difference (LSD) post hoc analysis revealed that in the Pre-test phase the game condition was significantly greater than video ($p < .05$) and classroom ($p < .05$) and the CBT condition was greater than video ($p < .01$), classroom ($p < .01$), and simulator ($p < .05$). In Phase 3 the game condition was greater than video ($p < .01$) and classroom ($p < .05$) and the CBT condition was greater than video ($p < .05$) and classroom ($p < .05$). Similarly, in Phase 4 the game condition was greater than video ($p < .01$) and classroom ($p < .01$) and the CBT condition was greater than video ($p < .01$) and classroom ($p < .01$). Similarly, in the post-test the game condition was greater than video ($p < .05$) and classroom ($p < .05$) and the CBT condition was greater than video ($p < .05$) and classroom ($p < .05$). In summary, during the pre-test, Phase 3, Phase 4, and the post-test the game and CBT conditions were greater than both the video and classroom conditions, and in the pre-test the CBT was also greater than the simulator condition.

The means for engagement by task phase and condition are presented in Table 2. There were significant differences in engagement in Phase 2 ($F = 5.05$, $p < .05$) and Phase 3 ($F = 8.64$, $p < .001$).

Table 2. Engagement means by phase and condition

	Pre-test	Phase 2	Phase 3	Phase 4	Post-test
Game	.355	.305	.320	.315	.382
Simulator	.457	.269	.336	.500	.470
CBT	.336	.116	.131	.261	.410
Video	.604	.665	.630	.690	.642
Classroom	.507	.408	.473	.464	.496

Fisher's LSD post hoc analysis revealed that in Phase 2 the video condition was greater than game ($p < .05$), simulator ($p < .01$), and CBT ($p < .01$). Additionally, the classroom condition was greater than CBT ($p < .05$). In Phase 3, video was greater than game ($p < .01$), simulator ($p < .01$), and CBT ($p < .01$). Also, the game condition ($p < .05$), simulator ($p < .05$), and classroom ($p < .05$) all had greater levels of distraction than CBT. Classroom had greater levels of distraction than game ($p < .05$).

The means for distraction by task phase and condition are presented in Table 3. However, there were no significant differences in distraction during any of the task phases.

Table 3. Distraction means by phase and condition

	Pre-test	Phase 2	Phase 3	Phase 4	Post-test
Game	.087	.129	.116	.106	.100
Simulator	.084	.154	.161	.151	.093
CBT	.001	.067	.029	.015	.010
Video	.014	.067	.046	.030	.018
Classroom	.128	.167	.133	.125	.070

The means for drowsiness by task phase and condition are presented in Table 4. There were significant differences in drowsiness in Phase 2 ($F = 4.11$, $p < .05$) and Phase 3 ($F = 14.56$, $p < .001$).

Table 4. Drowsiness means by phase and condition

	Pre-test	Phase 2	Phase 3	Phase 4	Post-test
Game	.001	.007	.006	.011	.030
Simulator	.007	.012	.007	.015	.013
CBT	.028	.037	.059	.032	.013
Video	.009	.001	.003	.001	.023
Classroom	.005	.015	.007	.012	.022

Fisher's LSD post hoc analysis revealed that in Phase 2 the CBT condition was greater than game ($p < .01$), simulator ($p < .05$), video ($p < .01$), and classroom ($p < .05$). Likewise, in Phase 3 the CBT condition was greater than game, simulator, video, and classroom (all $p < .001$). None of the other conditions significantly differed. from one another.

3.3 Performance and Learner Groups

The level of learning was calculated by looking at the difference between the pre- and post-tests. To avoid penalizing participants with strong prior knowledge who scored well on the pre-test, the amount of learning was calculated based on a percentage of possible learning.

$$Percent\ learned = \frac{(post\,test\ correct - pre\,test\ correct)}{(6 - pre\,test\ correct)}$$

Table 5. Performance scores by condition

	Pre-post change	Percent Learned
Game	1.25	.458
Simulator	2.50	.688
CBT	0.67	.278
Video	4.00	.875
Classroom	3.60	.910

The mean performance by subjects was examined by looking at both the change score between pre-test and post-test and by the percent learned score. The means of these measures is presented in Table 5. An ANOVA was conducted for both pre-post change (F = 6.36, p < .01) and percent learned (F = 3.37, p < .05), and both were significant. Post-hoc LSD analysis showed that the video condition was significantly greater than both the game and CBT, classroom was greater than game and CBT, and simulator was greater than CBT. Using percent learned as the performance measure, classroom is significantly greater than game and CBT, and video is greater than CBT. Therefore, in this population the best performance was obtained using either classroom or video.

The gauge values were correlated with the performance measures for percent learned. This was judged to be a more appropriate measure of learning because it didn't penalize participants who entered the training with some prior knowledge. The correlation between the percent learned performance measure and the gauge values through different ask phases is presented in Table 6. The trend is that workload is negatively related to performance, engagement is positively related, and drowsiness has a mixed relation depending on the task phase.

Table 6. Pearson correlation (r) values relating percent learned to gauge values

	Workload	Engagement	Distraction	Drowsiness
Pre-test	$-.431^{\wedge}$	NS	NS	NS
Phase 2	NS	NS	NS	NS
Phase 3	$-.480^{*}$	$.419^{\wedge}$	NS	$-.472^{*}$
Phase 4	$-.530^{*}$	NS	NS	NS
Post-test	NS	NS	NS	$.401^{\wedge}$

$^{\wedge}$ p < .10, * p < .05, NS = not significant.

Participants were divided into two groups based on the degree to which they learned during the task. A mean split was conducted to create high and low learner groups based on the percent learned calculation, and these means are presented in Table 7. These groups' means were then compared across different gauge values, and due to the small sample size (N = 7 and 11) and relatively large standard deviation none of the differences were statistically different although the means of pre-test drowsiness ($t = 1.86$, $p = .09$) were marginally different.

Table 7. Table of gauge means by learning rate and phase

	Workload		Distraction	
	Low learners	High learners	Low learners	High learners
Pre-test	.639	.586	.069	.082
Phase 2	.593	.591	.140	.111
Phase 3	.635	.595	.114	.099
Phase 4	.644	.591	.087	.114
Post-test	.659	.626	.068	.063

	Engagement		Drowsiness	
	Low learners	High learners	Low learners	High learners
Pre-test	.479	.407	.014	.003
Phase 2	.324	.352	.018	.012
Phase 3	.332	.427	.022	.003
Phase 4	.427	.435	.017	.010
Post-test	.469	.464	.017	.026

4 Discussion

The results of this study suggest that there are observable differences between participants as they participate in one of five training conditions. These differences exist across phases within the training as well as between the different training types (e.g., simulator, CBT, etc.). One of the most interesting findings is that, for this population sample, the best performance was observed in the classroom and video conditions. Individuals who participated in these conditions also experienced the lowest levels of cognitive workload during the training and had the highest levels of engagement. Conversely, the poorest performance was obtained in the CBT and game conditions, and individuals in these conditions experienced the highest levels of cognitive workload, some of the lowest levels of engagement, and CBT had the highest levels of drowsiness. The simulator condition was moderate on both performance and gauge values and didn't appear to trend with classroom and video or CBT and game.

The variation in gauge values throughout the training phases was examined, and the results indicate that the trend was for workload to start modestly, drop lower for Phase 2, increase towards initial levels in Phases 3 and 4, and climb to its peak in the post-test. This suggests that the post-test has the highest levels of workload, and this is in accordance with one's expectations that the testing phases of the training should be the most difficult. The trend for engagement was to follow a similar pattern to workload and drop during Phase 2, which was the initial presentation of the training material. Conversely, distraction was highest during Phase 2. This pattern for engagement and distraction is concerning, and it would be optimal to keep participant's engagement high and distraction low during Phase 2. Subjects in the video and classroom conditions who maintained high levels of engagement appeared to benefit from that in their performance. The trend for drowsiness is influenced by the high drowsiness scores of CBT participants in Phase 2 and 3, and this increase in drowsiness was likely associated with decreased performance.

Because of the nature of the study, it's not possible to determine whether the variation in mental state gauges is caused by the learning conditions. Also, it may be the case that the small number of participants in the CBT condition also happened to be the most drowsy. The value of the gauge data is in helping to determine whether a participant is maximizing the training experience. Regardless of whether it's the training condition or a participant's predisposition, if engagement is decreasing and drowsiness is increasing then the trainer should be informed. In the case of automated training, this information could be used to create an automated system response to decrease drowsiness, increase engagement, and likely decrease workload in the process. An example mitigation might be to replay the training sequence, give the participant a break to refocus, or present an engaging secondary task to reinvigorate the participant. Since the gauge values are calculated in real time, they could be integrated with the training software to provide a comprehensive solution.

Knowledge and skill acquisition is a complex cognitive process, and the efficacy of particular training methods should be based on both the post-test performance scores as well as the level of retention and real-world applicability, which could be verified through followup investigation of participants' skills. The results of this research have shown that, if one uses post-test performance as a guideline, monitoring participants' mental state variation provides valuable information to a trainer above and beyond task performance. This information has the potential to improve the training itself, regardless of training type, by monitoring and helping keep mental states optimal for learning. The current results echo the findings of Patten et al. [7] in that the better performers used less mental effort (and had lower cognitive workload). Furthermore, the current study expands the findings of Fournier et al. [8] and demonstrates that neurophysiological data can be valuable in a training environment, even with relatively modest training lengths of approximately 20 minutes. Further research will need to continue to clarify the impact that neurophysiological data can have on real-time monitoring to improve training as well as the development of training methods. The results of the current study suggest that a viable option exists in the former approach and raises expectations that the latter is feasible as well.

In conclusion, the evidence provided in this study suggests that the mental state gauges can provide predictive assessment of training outcome prior to completion of the training. The levels of workload and engagement that were observed throughout the training phases were indicators of the eventual outcome. This is a powerful concept, particularly when dealing with lengthy training regimes where there is a cost for repeating the entire training. Rather than awaiting the post-test assessment scores or performing multiple intrusive assessments during training, our results suggest that training efficacy could be unintrusively monitored during the training task and allow for focused and targeted intervention when workload and engagement deviate from ideal levels. Task-dependent mitigations could intervene and regain a participant's full attention or help reduce workload levels such that they are optimal, and then reset the training program to the time just before mental state levels became non-optimal.

In small classroom settings an instructor can pick up on cues from students and repeat information when it is clear that the students were distracted, drowsy, etc. Computer-based instruction doesn't automatically make these sorts of adjustments for its students. Including mental state gauges as part of a computer-based training solution would allow the system to make the individual adjustments necessary at distinct points in the training without waiting until the training has concluded and poor post-test performance indicates that they likely lost focus at some point during training. Maintaining constant awareness of trainee's mental state would allow a training program to sequentially present new material that builds upon previously presented material and to have the confidence that trainees maintained optimal training levels throughout. The next logical research step would be to develop a training system that presented a moderately lengthy training program, and to include an evaluation of training mitigations as part of the system. Concurrently, an in-depth investigation of the individual variation in mental state gauges and developing methods for consistent identification of non-optimal mental state levels would be required. Together, these research efforts would advance the understanding of how mental state gauges benefit training efficacy.

References

1. Mulert, C., Leicht, G., Pogarell, O., Mergl, R., Karch, S., Juckel, G., Möller, H.-J., Hegerl, U.: Auditory cortex and anterior cingulate cortex sources of the early evoked gamma-band response: Relationship to task difficulty and mental effort. Neuropsychologia 45, 2294–2306 (2007)
2. Gerlic, I., Jaušovec, N.: Differences in EEG power and coherence measures related to the type of presentation: Text versus multimedia. Journal of Educational Computing Research 25, 177–195 (2001)
3. Just, M.A., Carpenter, P.A., Miyaki, A.: Neuroindices of cognitive workload: neuroimaging, pupillometric and event-related potential studies of brain work. Theoretical Issues in Ergonomics Science 4, 56–88 (2003)
4. Tremoulet, P.D., Barton, J., Craven, P.L., Stibler, K., Wilcox, S., Gifford, A., Regli, S.H.: Disruptive Technologies Opportunities Fund (DTOF) Final Report, p. 32. Lockheed Martin ATL, Cherry Hill (2006)
5. Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P.D., Craven, P.L.: EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. Aviation, Space, and Environmental Medicine, Special Issue: Operational applications of cognitive performance enhancement technologies 78, B231–B244 (2007)
6. Tremoulet, P.D., Barton, J., Craven, P.L., Stibler, K., Wilcox, S., Gifford, A., Regli, S.H.: Service Test and Evaluation Report, p. 34. Lockheed Martin ATL, Cherry Hill (2006)
7. Patten, C.J.D., Kircher, A., Ostlund, J., Nilsson, L., Svenson, O.: Driver experience and cognitive workload in different traffic environments. Accident Analysis & Prevention 38, 887–894 (2006)
8. Fournier, L.R., Wilson, G.F., Swain, C.R.: Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: Manipulations of task difficulty and training. International Journal of Psychophysiology 31, 129–145 (1999)

Identifying the Nature of Knowledge Using the Pressures Applied to a Computer Mouse

Martha E. Crosby[1], Curtis Ikehara[1], and Wendy Ark[2]

[1] University of Hawaii at Manoa, 1680 East-West Road, Honolulu, HI 96822, USA
[2] IBM Almaden Research Center, 650 Harry Rd. San Jose, CA 95120, USA
{Crosby,Cikehara}@hawaii.edu, wsark@us.ibm.com

Abstract. The nature of knowledge retention is not that a student either knows or doesn't. Using signal detection theory, the correct and incorrect responses a student provides can each be subdivided into two more levels of knowledge using the student's confidence of answer correctness. The proposed study will attempt to link confidence of answer correctness to the categorized pressures applied to a computer mouse allowing for the partitioning of responses. Twenty participants that were part of a pedagogical methods study will be retested using a computer-based multiple choice test and pressure sensitive computer mouse. Participants will also rate their confidence of answer correctness. It is hypothesized that the analyzed pressures applied to the computer mouse will indicate the confidence of answer correctness. Using the categorized pressures from the computer mouse allows for the real-time assessment of a student's knowledge to guide pedagogical follow-up.

Keywords: knowledge**,** pressure sensitive computer mouse, confidence of correctness.

1 Introduction

Assessing how much students truly understand is an age-old challenge for educators. An ideal assessment is one that can easily probe a student on materials, but can understand the depth of knowledge about the materials. A good instructor can easily question a student to investigate whether they have a complete understanding or if they still lack basic knowledge; however, these types of assessments are not always feasible. Many educators rely on multiple choice exams, which have many benefits such as ease of grading, standardization, and scalability. However, how accurate can multiple choice exams be for assessing the knowledge of materials? When a student answers a multiple choice question correctly, it is assumed that the student has sufficient knowledge about the topic of interest. Should the student answer incorrectly, it is assumed that some form of review is required. However, the answer given, particularly on a multiple choice exam, does not give transparency to whether the student actually understood the question or the topic.

The nature of knowledge retention is not that a student either knows or doesn't, but is much more complex. Using signal detection theory [1], both correct and incorrect answers can be subdivided into two categories.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 595–600, 2009.

1. Correct
   a. Based on appropriate knowledge (i.e., Desired)
   b. Correct – Based on guessing (i.e., Lack of knowledge)
2. Incorrect
   a. Lack of knowledge
   b. Based on inappropriate knowledge

Were there a way to identify these four cases, the follow-up to the response could be more appropriate. Case 1a is recognized as the desired educational outcome and the student would be rewarded for being correct. Case 1b and 2a would both require review of the topic to increase the student's knowledge and Case 2b would require correction of the inappropriate knowledge. Without being able to identify the cases, Case 1b would not require the needed topic review and Case 2b would not specifically address identifying and correcting the inappropriate knowledge.

Detecting these four cases is important to apply the correct educational remediation. As described above, not identifying these cases leads to inappropriate education follow-up half the time. Identifying guessing behavior is possible by measuring the response time based on previously difficulty rated questions from a large number of subjects [2], but this is impractical for regular and small size classes and would require always using the same difficulty rated test questions.

Self-assessment of confidence of judgment of the correct answer is a possible method to identifying the four cases. People generally tend to be over-confident about answer correctness, but there is a correlation between confidence and answer correctness [3]. The feedback on answer correctness the student sees when viewing the results of the post lesson test will tend to better calibrate and reduce the overconfidence of judgment of answer correctness [4]. The problem with the self-assessment of confidence of answer correctness is that it is a disruptive burden for the student during testing.

This paper proposes a more automated system for identifying the four cases. Preliminary pilot studies using the click signature obtained from pressure sensors on a computer mouse acquired during the task performance has indicated that a greater distortion in the signature occurs when the task increases in difficulty. It has also been shown that the pressures applied to a computer mouse can be use to assess cognitive load [5, 6, 7, 8]. It is hypothesized that the mouse click signature can be used to identify the four cases that occur when responding to a question.

Among the questions this study will attempt to address are:

1. Is confidence of judgment a good indicator of correctness of answer?
2. Is the mouse click signature a good indicator of correctness of answer?
3. Does confidence of judgment correlate with mouse click signatures?

Table 1 shows in bold the two desired response paths for a student who has knowledge of a topic and has answered a question in a post lesson test upon completing a topic. Then a few months later answers the same or similar question in a retention test. The first desired path of the state of knowledge is that the student retained the correct knowledge to respond to a question. The state of knowledge should be correlated to the student's confidence rating and in turn also correlated to the categorized pressures applied to a computer mouse. The second desired path of the state of

Table 1. Response to the same or similar question in post lesson and retention tests

Post lesson test response	Retention test response	State of Knowledge	Student's confidence rating	Categorized Mouse Pressure
Correct	**Correct**	**Retained correct knowledge of question (1a)**	**Certain**	**Lower distortion**
		Initially guessed correct in post-test (1b)	Uncertain	Higher distortion
	Incorrect	Did not retain correct knowledge (2a)	Uncertain	Higher distortion
		Incorrectly retained knowledge (2b)	Certain	Lower distortion
Incorrect	**Correct**	**Corrected the state of knowledge (1a)**	**Certain**	**Lower distortion**
		Correct guess in retention test (1b)	Uncertain	Higher distortion
	Incorrect	Did not retain correct knowledge (2a)	Uncertain	Higher distortion
		Retained Incorrect knowledge (2b)	Certain	Lower distortion

knowledge is when a student was incorrect in the post lesson test, but corrected and retained the correct state of knowledge. The student would be similarly confident of answer correctness and have a correlated mouse pressure response.

The words marked in bold are the two desired response paths for a student answering a question in the retention test indicating the state of knowledge, confidence rating and categorized mouse pressure.

2 Method

Approximately twenty participants will be retested using a computer-based multiple choice survey and pressure sensitive computer mouse. Participants will also rate their confidence of answering correctly.

2.1 Subjects

Approximately twenty participants that were in a CPATH grant to assess different methods of teaching computer science will be recruited for retesting of the knowledge they obtained in the computer class. The class would have ended two to four months ago and this test will access the participant's retention of the content of the class. Data of answers to the same or similar questions will be available for analysis.

2.2 Equipment

Participants will be taking the retention test using a computer and respond to questions by only using the computer mouse. The computer mouse being used is a

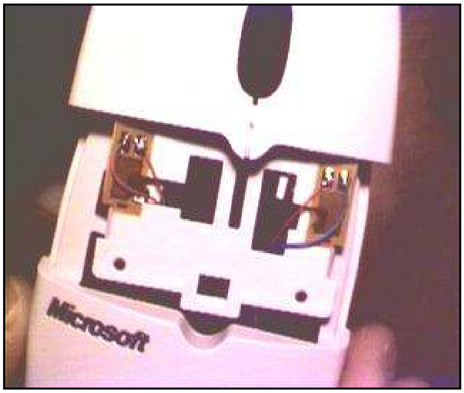


Fig. 1. Internal view of a computer mouse with pressure sensors on the buttons

pressure sensitive computer mouse (PSCM) where data on the pressures applied to the PSCM will be collected for analysis and categorization (see Figure 1).

2.3 Task

The task will be a set of computer based multiple choice questions. The first set will consist of "confidence of answer correctness" calibration questions. Then a set of topic retention questions from the previously taken computer programming class will be given followed by a final set of calibration questions.

Calibration Questions: The set of calibration questions will consist of ten multiple choice response questions selected to create varying levels of "confidence of answer correctness." An example of a question that should elicit a high "confidence of answer correctness" would be: "What is the sum of 4 + 9 + 10?" An example of a question that should elicit a low level of confidence of answer correctness would be: "What is the remainder of 37871 divided by 97?" Participants will be asked on the subsequent question to rate "confidence of answer correctness" for each question.

Confidence of Answer Correctness Question: A standard 5-level Likert scale will be used to assess the participant's "confidence of answer correctness." An example is shown below.

> The answer I gave to the previous question is correct. Please select one response.
>
> Strongly disagree
> Disagree
> Neither agree or disagree
> Agree
> Strongly Agree

Retention Questions: The retention test will be comprised of five to ten multiple choice response questions that are similar or the same as those the student had been previously tested on in the post lesson test. Participants will be asked on the subsequent question to rate "confidence of answer correctness" for each question. An example of a retention question is shown below.

> What will be the value of i after the statements at the right have been executed?
>
> | | | |
> |---|---|---|
> | A. | 11 | int i = 10; |
> | B. | 33 | int j = 33; |
> | C. | 34 | while ((3 * ++i) < j) |
> | D. | 10 | i = i + j++; |
> | E. | None of the above | |

2.4 Procedure

The participant will be briefed on the three sets of questions (i.e., calibration, confidence & retention) and how to answer the questions. There will be a time limit for each set of questions. The participant will not be allowed to use any aids when determining the answers to the questions.

3 Results

Analysis 1: An analysis on the data collected from the pressure sensitive computer mouse (PSCM) during the calibration question will compare high to low "confidence of answer correctness" to the pressures applied to the computer mouse. Previous pilot studies indicate that minimal difficulty is indicated by a sudden sharp pressure on the mouse button when selecting a response while high difficulty is indicated by a distortion of the pressure on the mouse button. It is hypothesized that responses to the "confidence of answer correctness" will be correlated to the level of distortion of the pressures applied to the computer mouse.

Analysis 2: An analysis on the data collected from the pressure sensitive computer mouse (PSCM) during the retention question will compare high to low "confidence of answer correctness" to the pressures applied to the computer mouse. It is hypothesized that to obtain the optimal correlation between "confidence of answer correctness" and the categorization based on the level of distortion of the pressures applied to the computer mouse for the set of retention questions, the normal distortion for an individual in mouse pressures as determined from Analysis 1 will be needed to normalize the data for each individual. Previous studies have found individual differences in pressure variation unique to the individual [9, 10].

4 Discussion

Assuming Analysis 1 and Analysis 2 can produce a reliable categorization of the pressures applied to the computer mouse indicative of the "confidence of answer correctness" it becomes possible to distinguish with the pressure sensitive computer mouse in real-time the four cases of response to a question. The benefit of real-time assessment of a student's knowledge to a question into four cases is that appropriate pedagogical follow-up can be executed immediately following the response. Using this method can improve both the assessment of a student's knowledge and appropriately respond to the state of a student's knowledge improving the learning outcome.

References

1. Swets, J.A.: Signal detection and recognition by human observers. In: Swets, J.A. (ed.) Contemporary readings, Wiley, Chichester (1964)
2. Yang, X.: Methods of Identifying Individual Guessers From Item Response Data. Educational and Psychological Measurement 67, 745 (2007)
3. Harvey, N.: Confidence in Judgment. Trends in Cognitive Sciences 1(2), 78–82 (1997)
4. Gonza´lez-Vallejo, C., Bonham, A.: Aligning Confidence with Accuracy: Revisiting the Role of Feedback. Acta Psychologica 125, 221–239 (2007)
5. Ikehara, C.S., Crosby, M.E.: Using Real-Time Physiological Monitoring for Assessing Cognitive States. In: Ghinea, G., Chen, S. (eds.) Digital Multimedia Perception and Design, ch. 8, pp. 170–185 (2006)
6. Crosby, M.E., Ikehara, C.S.: Using Physiological Measures to Identify Individual Differences in Response to Task Attributes. In: Augmented Cognition International Conference, San Francisco, CA (2006)
7. Ikehara, C.S., Crosby, M.E., Chin, D.N.: A Suite of Physiological Sensors for Assessing Cognitive States. In: 11th International Conference on Human-Computer Interaction (July 2005)
8. Ikehara, C.S., Crosby, M.E.: Assessing Cognitive Load with Physiological Sensors. In: 38th Annual Hawaii International Conference on System Sciences (January 2005)
9. Crosby, M.E., Ikehara, C.S.: Biometric Technology for Human Identification Continuous Identity Authentication Using Multimodal Physiological Sensors. In: SPIE Defense and Security Symposium (April 2004)
10. Ikehara, C., Crosby, M.E.: User Identification Based on the Analysis of the Forces Applied by a User to a Computer Mouse. In: Proceedings of the Hawaii International Conference on System Sciences, Kona, Hawaii (January 2003)

Realizing Adaptive Instruction (Ad-In): The Convergence of Learning, Instruction, and Assessment

Edward Dieterle and John Murray

SRI International
333 Ravenswood Avenue, Menlo Park, CA 94025-3493 USA
edward.dieterle@sri.com, john.murray@sri.com

Abstract. In this paper, we define adaptive instruction, or Ad-In, as applied to sophisticated skills development systems that target learning and assessment in a highly individualized and interactive manner. We argue that the successful design and use of such systems rely heavily upon the interrelationships among learning styles, instructional theories, and assessment methods, in the context of personalized learning. We outline and structure the links among these topics by drawing upon recent empirical studies of virtual environments and augmented realities. The paper also presents a candidate architecture for applying Ad-In concepts in an intelligent interactive environment for skills development.

Keywords: Adaptive instruction, augmented reality, immersion, intelligent tutoring assistant, multi-user virtual environment, neomillennial learning styles.

1 Introduction

Adaptive Instruction (Ad-In), as argued in this paper, is characterized by dynamic programs of instruction that are readily capable of adapting or of being adapted to individual learning requirements. Ad-In addresses the unique and situation-specific needs of learners by concurrently providing clear information, opportunities for thoughtful practice, informative feedback, and a favorable combination of intrinsic and extrinsic motivators tailored to the individual learner. Central to Ad-In are interactive and immersive technologies that target learning, instruction, and assessment. Such learning innovations build on previous research, which demonstrates that technology enhances learner understanding when it (a) supports learning in real-world contexts, (b) connects learners to experts and communities of learners, (c) makes possible visualization and analysis tools for thinking with data and datasets, (d) scaffolds problem solving that enables more complex reasoning than possible otherwise, and (e) provides opportunities for feedback, reflection, and revision of knowledge construction [1].

Delving into and articulating Ad-In, we examine the interrelationships among learning styles, instructional theories, and assessment methods, in the context of personalized learning. We draw upon recent experience in empirical studies of virtual environments and augmented realities, and we explore how new learning styles, instructional theories, and tools for measuring understanding are emerging from the use

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 601–610, 2009.

of these technologies, which may affect training program development and delivery from an adaptive instruction perspective.

2 Media-Based Learning Styles

Learning styles are theoretical constructs designed to help explain the learning process, a complex and nuanced phenomenon. Learning styles are comprised of (a) *cognitive styles*, which consider concept formation and retention and sensory reception; (b) *affective styles*, which consider attention, expectancy, and incentive; and (c) *physiological styles*, which consider the functions and activities of human organisms, including all physical and chemical processes [2].

Early scholars and researchers of learning styles held as axiom that styles were inflexible, context-independent, and solely determined by ability and personality [3]. Modern conceptualizations of styles reject such principles and view the construct as (a) shaped by physical and mental development, personal interests, and sociocultural influences; (b) preferences in the use of abilities, not abilities themselves; (c) existing within all people in varying degrees, resulting in profiles of styles; (d) variable across tasks and situations, having the potential to change over time; (e) measurable, teachable, and socializable; and (f) variable in terms of flexibility and adaptability within people [4-6].

A growing number of researchers and scholars have begun investigating media-based learning styles, which are modern learning styles understood in relation to three complementary human-computer interfaces that are reshaping thinking, learning, and instruction [7]. The *World-to-the-Desktop*, the first and most mature interface, is facilitated through laptop, desktop, and tablet computers connected to the Internet. By bringing the world to the user, this interface provides users access to archives and sophisticated databases and also enables collaborations, mentoring relationships, and virtual communities-of-practice [8, 9]. *Multi-User Virtual Environments* (MUVEs), the second interface, are characterized by participants controlling digital emissaries to engage digital content and interact with fellow users to complete various kinds of tasks in three-dimensional virtual contexts. At a time when nine of the ten best selling computer games of 2007 are MUVEs, a growing number of projects have developed MUVEs specifically for teaching and training [10, 11]. *Ubiquitous computing*, a third human-computer interface, provides dynamic, temporal, and contextually specific tools through computers that are no longer perceptually foregrounded [12, 13]. Interactivity seamlessly and imperceptibly integrates into activity. On a variety of scales, users obtain ever-present connectivity and access to capture, process, send, and receive information through multiple devices anytime and anywhere [14-16]. Participatory simulations and augmented realities (ARs) made possible through wireless handheld computers have provided the basis for learning and teaching using ubiquitous computing [17-19].

One primary difference among the World-to-the-Desktop interface, MUVEs, and ARs is immersion. Immersion can induce a user into a perceived state of being present with others or in a place other than where the user is physically located; it depends in part on the ability to empower actions and activity while facilitating affective factors that influence learning, such as emotional awareness, self-control, and self-efficacy [20, 21]. Given that the World-to-the-Desktop interface is context independent, it cannot bring about a sense of "being there" to the same extent that MUVEs and ubiquitous computing can support a sense of "presence" in a virtual world or an AR. Such differences make possible learning in MUVEs and ARs that supports the situational and distributional nature of cognition with respect to thinking, learning, and doing in ways that are limited or absent in World-to-the-Desktop computer interactions.

As an illustrative example contrasting learning through an immersive versus non-immersive interface, a learner studying disease and disease transmission with the World-to-the-Desktop might communicate with epidemiologists via email or join a listserv devoted to the transmission and control of epidemic diseases, thus beginning an ongoing exchange of ideas and questions. The ebb and flow of information through the World-to-the-Desktop interface, however, is not generally characterized as immersive. The learner is not part of the events he or she is studying. Rather, he or she is a distant observer as compared to an active participant. The River City Project, on the other hand, uses a MUVE called "River City" to support the situated study of disease and disease transmission [22, 23]. Based on authentic geographical, historical, and sociological conditions, River City is a town besieged with health problems that affect the wellbeing of its residents. The mayor of River City has commissioned learners to travel back in time, bringing their 21st-century knowledge and technology to address a 19th-century epidemic. The affordances of the MUVE and its accompanying storyline allow learners to think and act as scientists in an environment of intermediate complexity. It is less complex than the real world, which can be overwhelming, but more complex, authentic, and nuanced than a "cookbook" lab, which is designed to be instructor- and learner-proof. If asked where learners are located while interacting with the River City MUVE, users are likely to state they are in River City and with their teammates instead of where they are located physically.

As a second illustrative example, "Reliving the Revolution" (RtR) uses wireless handheld devices to support an AR game that teaches historic inquiry, effective collaboration, media fluency, decision-making, and critical thinking skills [24]. RtR enables participants to traverse the present-day site of the Battle of Lexington to relive this historic battle from the American Revolution through the eyes of one of four historic figures. Participants use their device to collect information or evidence to determine who fired the first shot in the Battle, a source of continued debate in American history. GPS-enabled devices provide participants location-based virtual information on the social, historical, economic, geographic, and political processes relevant to both the Battle of Lexington and the American Revolution.

River City and RtR utilize key aspects of Ad-In and are readily capable of (a) adapting or of being adapted to individual learning requirements; (b) addressing the unique and situation-specific needs of learners by concurrently providing clear information, opportunities for thoughtful practice, informative feedback, and a favorable combination of intrinsic and extrinsic motivators tailored to the individual learner; and (c) utilizing interactive and immersive technologies that target learning, instruction, and assessment [22, 25].

3 Neomillennial Learning Styles

The relations between participatory and immersive media and learning styles, shown in Figure 1, have become an important new research direction. Research on MUVEs, ARs, and other immersive, personalized, and interactive media led Dede and colleagues to propose a new classification of media-based learning styles [26].

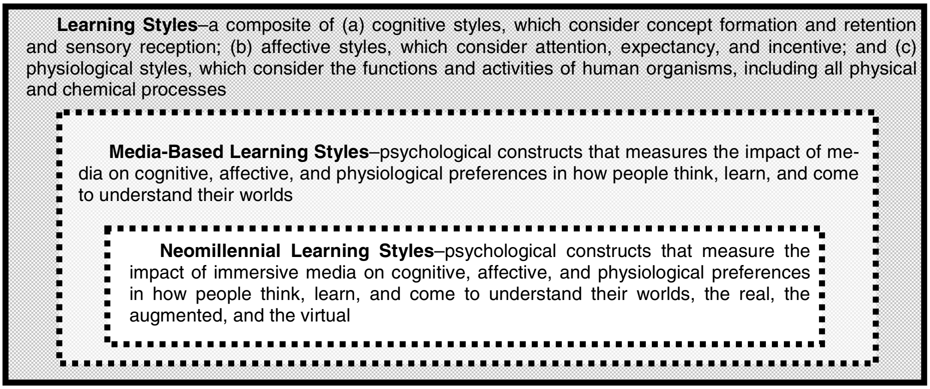


Fig. 1. Nested diagram depicting the interconnections between learning styles, media-based learning styles, and neomillennial learning styles

"Neomillennial" Learning Styles (NLS) include a person's preferred cognitive, affective, and physiological styles in which they think, learn, and come to understand their worlds, in the real, augmented, and virtual domains and are characterized by:

- Fluency in multiple media, valuing each for the types of communication, activities, experiences, and expressions it empowers.
- Learning based on collectively seeking, sieving, and synthesizing experiences rather than individually locating and absorbing information from some single best source; preferring communal learning in diverse, tacit, situated experiences; valuing knowledge distributed across a community and a context, as well as within an individual.
- Active learning based on experience (real and simulated) that includes frequent opportunities for embedded reflection; valuing bicentric, immersive frames of reference that infuse guidance and reflection into learning-by-doing.
- Expression through nonlinear, associational webs of representations rather than linear stories (for example, authoring a simulation and a Web page to express understanding rather than writing a paper); using representations involving richly associated, situated simulations.
- Co-design of learning experiences personalized to individual needs and preferences.

NLS are present in varying degrees in learners of all ages and not just "digital natives" [7]. Ongoing interaction with immersive technologies, such as MUVEs and ARs, develop and enhance NLS. Dieterle and colleagues have utilized both qualitative and quantitative methods to study NLS in MUVEs and ARs [25, 27, 28]. Such studies of the links between MUVEs and ARs and learning theory have produced valuable insights for designing more effective ways to adapt instructional processes to the learning style of the learner. For example, learners who generally enjoy tasks that require creative strategies, such as working with ideas in new ways, and mashing up and sharing information, appear to be more well suited for learning scientific problem solving skills in MUVEs than those who avoid the same activities and don't share the same predilections [27].

4 Instructional Theories

Previous generations of instructional design tended to provide all learners with uniform experiences that required learners to adapt to the pedagogy [29]. Cost savings from the systemization of schools, resulting from the mass production and distribution of materials and techniques, were primary factors motivating the use of the factory model of instruction in many formal learning institutions [30]. This instructional philosophy, however, conflicts with contemporary research into how people learn, revealing that, with enough time, access, guidance, and motivation, almost everyone can learn just about anything to a great extent and yet, almost no one learns exactly the same way, through the same pathways, or to the same degree [31-34].

Where modern instructional theories advocate for personalized instruction, an inability to leverage an economy of scale has limited and restricted the scope and freedom to implement personalized instruction widely. Efforts of researchers and engineers, however, are on the verge of changing the way instruction is personalized and adapted to individual learners profoundly. Applying scientific knowledge of mind, brain, and education to generate economically viable solutions that address the challenges associated with advancing personalized learning significantly to large numbers of individuals is one of the National Academy of Engineers 14 grand challenges [35].

Ad-In involves orchestration among members of the research team and participating instructors and learners, which can be understood through a music metaphor with the research team as composers, instructors as conductors, and learners as musicians. All three groups work in harmony to co-design learning experiences that are personalized to individual needs and preferences, while adhering to the spirit of the curriculum. As composers, the research team develops a curriculum. Instructors, in turn, receive the curriculum and act as conductors, using knowledge of the local culture and learning setting to getting the most out of their learners. The instructor’s role is to guide learners’ performance through immersive and interactive experiences. Diversity of prior knowledge among learners provides a wealth of experience and knowledge from which teams can draw upon to engage the complexities and challenges the curriculum provides. Just as musicians tend to specialize, not all learners need to master every aspect of the curriculum equally well. Instead, teammates play off each other’s strengths while buoying up their collective weaknesses to produce the best team performance possible.

Increased levels of challenge, incremental growth of understanding, and ongoing opportunities for success characterize adaptive instruction. The focus of the learning environment is the learner, rather than the content or the instructor. Learners are not viewed as blank slates upon entry into the learning environment. One-size does not fit the needs, skill levels, interests, or abilities of all learners. Cultural differences and prior knowledge add to variation among learners. Learners work through important and relevant content topics cohesively (as compared to piecemeal) that encourage doing with understanding (as apposed to simply hands-on doing). Learners are regularly given assessment opportunities to demonstrate what they know (and don't know) and can do (and cannot do) within learning activities (as compared to stepping out of the activity to complete an assessment). Learners and their instructors use the formative assessment feedback to understand the learner's progress and to shape and guide instruction. We should carefully distinguish this aspect of learner assessment from the process of testing if the learner can be certified as having 'passed the course'.

5 Assessment Methods and Tools

We recognize the need for methods of assessing the impacts that the technology (i.e. the adaptive instruction and automated tutoring techniques) has upon the learner and the learning experience. This is the traditional domain of researchers' experimental data gathering and analysis of outcomes. As Sheingold and Frederiksen observe, "to change our expectations about what students should know and be able to do will involve also changing both the standards by which student achievements are judged and the methods by which student's accomplishments are assessed" [36].

An insightful new theoretical frame put forth by Solomon and Perkins identified three levels by which technology influences thinking and learning. The immediate results are the *effects with* a technology, which resulting in expanded cognitive capacity and amplified perception. After considerable experience with a technology, users exhibit the *effects from* a technology, the residual impact of a technology when it is no longer present. The most profound effects are the *effects through* a technology, which fundamentally reorganize cognitive activity [37]. Accurately assessing effects with, from, and through technology requires measurement methodologies and objectives that match our evolving expectations for learning outcomes, as well as new ways in which they learn.

Educational MUVEs such as River City and intelligent tutoring systems (ITSs) have the ability to record and store every keystroke users generate inside the MUVE or with the tutor [22, 38]. Through such technologies, researchers can collect, store, retrieve, process, and analyze information on the activities of individual users, teams, or groups of teams as they participate in the simulation. The level of detail in these records is comprehensive, indicating exactly where students went, with whom they communicated, what virtual artifacts they activated, and how long each of these activities took. This richly varied store of data can couple with other artifacts of learning to develop novel, performance-based assessments of complex performances that leverage NLS, disciplinary reasoning, and procedural skills.

6 Adaptive Instruction in Context

We now briefly examine how Ad-In can be applied within the context of an intelligent interactive environment for individualized skills development. Figure 2

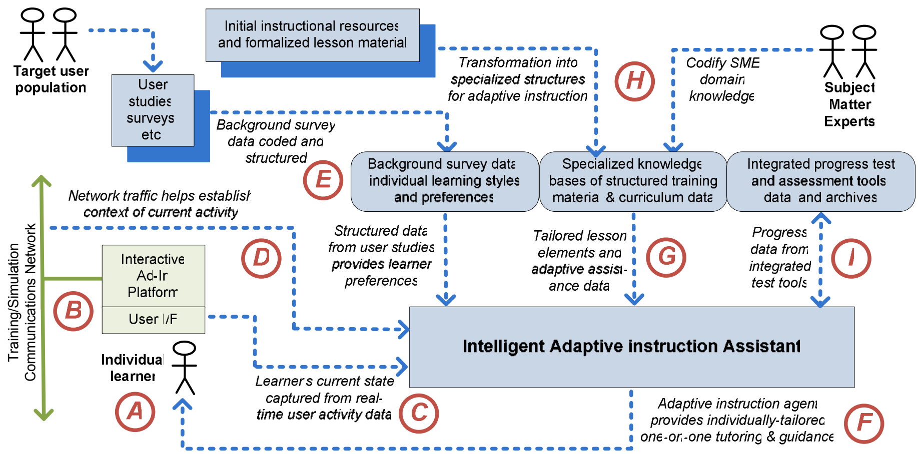


Fig. 2. Integration of an intelligent Ad-In assistant into an interactive, adaptive instructional environment

illustrates how an automated Ad-In tutoring assistant may be used in conjunction with a target application system. This example is structured around a typical military training system, although the concepts discussed are broadly applicable in other application domains.

A human learner interacts with the target application via the training platform (A), which may be a standalone unit or, quite frequently, linked to a broader networked simulation/training infrastructure (B). The user interface of the platform is instrumented so that data and observations on the learner's activity are sensed and provided to the Ad-In assistant in real-time (C). Such sensors may include augmented cognition tools, such as eye trackers or EEG devices, as well as more traditional user interaction monitors. The assistant also monitors the communications network, possibly both voice and data traffic, which helps to situate the context of the learner's current activities and goals (D). Information from background user studies and individual survey data (E) is available, which the assistant can draw upon to help identify the learner's likely behavioral preferences and learning style.

These inputs enable the Ad-In system to react to changes observed in the training activity, and to provide individualized tutoring support, dynamically and inconspicuously, as the lesson proceeds (F). To do this, it draws upon knowledge bases and teaching plans that have been specially structured for use in the Ad-In environment (G), and which have been constructed from instructional material, curriculum data, and reference manuals, as well as broader background knowledge sources, including subject matter experts (H). The assistant also retains and uses the learner's prior lesson history, as well as data from integrated progress tests and assessments, to document the learner's areas of improvement, and to identify elements where additional work is needed, thus enabling the system to adapt the instructional process to the evolving requirements of the individual (I).

7 Conclusions

We believe that the Ad-In approach, in which the independent but interrelated elements of learning styles, instructional theories, and assessment methods comprise a unified whole, is a particularly useful way to characterize adaptive instruction systems in the realm of augmented cognition. Although scientists that study how people learn have for some time believed that media impacts thinking, learning, and understanding, it is only relatively recently that we have found reliable ways of articulating media-based learning styles from empirical evidence. Thus far, quantitative and qualitative studies of this type have concentrated upon measuring the cognitive and affective preferences of the learner with regard to MUVEs and ARs. The addition of new data from refined studies of cognitive, affective, and physiological styles will offer valuable new insights into future strategies for adaptive instruction and augmented cognition. Further research into quantitative measures, which measure the 'what' and 'how', complimented by qualitative measures, which measure the 'why', can help to predict which learners should be directed toward immersive game-like training and which should be provided alternative learning experiences.

Acknowledgements

The authors wish to thank their colleagues from SRI International's Center for Technology in Learning and Engineering and Systems Division for their thoughtful feedback on early drafts of this paper.

References

1. National Research Council: How people learn: Brain, mind, experience, and school. National Academy Press, Washington, DC (2000)
2. Keefe, J.W.: Learning style: Theory and practice. National Association of Secondary School Principals, Reston (1987)
3. Sternberg, R.J., Grigorenko, E.L.: A capsule history of theory and research on styles. In: Sternberg, R.J., Zhang, L.-f. (eds.) Perspectives on thinking, learning, and cognitive styles, pp. 1–21. Laurence Erlbaum Associates, Mahwah (2001)
4. Zhang, L.-f., Sternberg, R.J.: The nature of intellectual styles. Lawrence Erlbaum Associates, Mahwah (2006)
5. Sternberg, R.J.: Thinking styles. Cambridge University Press, Cambridge (1997)
6. Sternberg, R.J., Zhang, L.-f. (eds.): Perspectives on thinking, learning, and cognitive styles. Laurence Erlbaum Associates, Mahwah (2001)
7. Dede, C., Dieterle, E., Clarke, J., Ketelhut, D.J., Nelson, B.: Media-based learning styles: Implications for distance education. In: Moore, M. (ed.) Handbook of distance education, pp. 339–352. Lawrence Erlbaum Associates, Mahwah (2007)
8. Ardichvili, A., Page, V., Wentling, T.: Motivation and barriers to participation in virtual knowledge-sharing communities of practice. Journal of Knowledge Management 7, 64–77 (2003)
9. Dubé, L., Bourhis, A., Jacob, R.: The impact of structuring characteristics on the launching of virtual communities of practice. Management 18, 145–166 (2005)

10. Entertainment Software Association: Essential facts about the computer and video game industry: 2008 sales, demographic, and usage data. Entertainment Software Association, Washington (2008)
11. Dieterle, E., Clarke, J.: Multi-user virtual environments for teaching and learning. In: Pagani, M. (ed.) Encyclopedia of multimedia technology and networking. Idea Group, Inc., Hershey (2008)
12. Weiser, M.: The computer for the twenty-first century. Scientific American 265, 94–104 (1991)
13. Abowd, G.D., Mynatt, E.D.: Charting past, present, and future research in ubiquitous computing. ACM Transactions on Computer-Human Interaction 7, 29–58 (2000)
14. Mitchell, W.J.: Me++: The cyborg self and the networked city. MIT Press, Cambridge (2003)
15. Rheingold, H.: Smart mobs: The next social revolution. Perseus Publishing, Cambridge (2002)
16. Greenfield, A.: Everyware: The dawning age of ubiquitous computing. New Riders, Berkeley (2006)
17. van 't Hooft, M., Swan, K. (eds.): Ubiquitous computing in education: Invisible technology, visible Impact. Lawrence Erlbaum Associates, Mahwah (2006)
18. Dieterle, E.: Wireless handheld devices as tools for thinking and learning in university settings. Harvard Graduate School of Education, Cambridge (2006)
19. Klopfer, E.: Augmented learning: Research and design of mobile educational games. MIT Press, Cambridge (2008)
20. Dede, C.: The evolution of constructivist learning environments: Immersion in distributed, virtual worlds. Educational Technology 35, 46–52 (1995)
21. Winn, W.: Learning in artificial environments: Embodiment, embeddedness and dynamic adaptation. Technology, Instruction, Cognition and Learning 1, 87–114 (2003)
22. Dieterle, E.: Games for science education. In: Hirumi, A. (ed.) Playing [video] games in schools: Engaging learners through interactive entertainment, International Society for Technology in Education, OR (in press)
23. Ketelhut, D.J., Dede, C., Clarke, J., Nelson, B., Bowman, C.: Studying situated learning in a multi-user virtual environment. In: Baker, E., Dickieson, J., Wulfeck, W., O'Neil, H. (eds.) Assessment of problem solving using simulations. Lawrence Erlbaum Associates, Mahwah (2007)
24. Schrier, K.: Reliving history with "reliving the revolution": Designing augmented reality games to teach the critical thinking of history. In: Gibson, D., Aldrich, C., Prensky, M. (eds.) Games and simulations in online learning: Research and development frameworks, pp. 250–270. Information Science Publishing, Hershey (2007)
25. Dieterle, E., Dede, C., Schrier, K.L.: "Neomillennial" learning styles propagated by wireless handheld devices. In: Lytras, M., Naeve, A. (eds.) Ubiquitous and pervasive knowledge and learning management: Semantics, social networking and new media to their full potential, pp. 35–66. Idea Group, Inc., Hershey (2007)
26. Dede, C.: Planning for "neomillennial" learning styles: Implications for investments in technology and faculty. In: Oblinger, J., Oblinger, D. (eds.) Educating the net generation, pp. 226–247. Educause Publishers, Boulder (2005)
27. Dieterle, E.: Neomillennial learning styles and River City. Children in Technological Environments: Interaction, Development, and Design, a Special Issue of Children, Youth and Environments 19 (in press)

28. Dieterle, E.: The convergence of neomillennial learning styles and a multi-user virtual environment. Teaching and Learning. Harvard Graduate School of Education, Cambridge (2008)
29. Darling-Hammond, L.: Futures of teaching in American education. Journal of Educational Change 1, 353–373 (2000)
30. Tyack, D.B.: The one best system: A history of American urban education. Harvard University Press, Cambridge (1974)
31. Kegan, R.: In over our heads: The mental demands of modern life. Harvard University Press, Cambridge (1994)
32. Gardner, H.: Multiple intelligences: New horizons. Basic Books, New York (2006)
33. Battro, A.M., Fischer, K.W., Léna, P.J. (eds.): The educated brain: Essays in neuroeducation. Cambridge University Press, Cambridge (2008)
34. Duckworth, E.R.: "Tell me more": Listening to learners explain. Teachers College Press, New York (2001)
35. National Academy of Engineering: Grand challenges for engineering, vol. 2009 (2008)
36. Sheingold, K., Frederiksen, J.: Using technology to support innovative assessment. In: Means, B. (ed.) Technology and education reform: The reality behind the promise, pp. 111–132. Jossey-Bass, San Francisco (1994)
37. Salomon, G., Perkins, D.: Do technologies make us smarter? Intellectual amplification with, of and through technology. In: Sternberg, R.J., Preiss, D. (eds.) Intelligence and technology: The impact of tools on the nature and development of human abilities, pp. 71–86. Lawrence Erlbaum Associates, Mahwah (2005)
38. Anderson, J.R., Greeno, J.G., Reder, L.M., Simon, H.A.: Perspectives on learning, thinking, and activity. Educational Researcher 29, 11–13 (2000)

Adaptive Learning via Social Cognitive Theory and Digital Cultural Ecosystems

Joseph W. Juhnke and Adam R. Kallish

Tanagram Partners, 125 North Halsted Street, Suite 400
Chicago, Illinois 60661, USA
{joseph.juhnke,adam.kallish}@tanagram.com

Abstract. This paper will look at the human predisposition to oral tradition and its effectiveness as a learning tool to convey mission-critical information. After exploring the effectiveness of the conveyance of information, the paper will examine current adaptive learning research and develop a system that will marry the strengths of oral tradition with those of an optimal adaptive learning environment. Emphasis will be made in the area of military service personnel stationed in contested cultures, the aiding of their arrival and once established their continual improvement processes. This paper will then illustrate a digital cultural ecosystem that leverages the strengths of current industry thinking in digital community development and social architecture combining the adaptive learning models discussed earlier to create a dynamic digital social ecology that could significantly improve the transition process by exposing service personnel to the collective learning of all of the personnel currently and previously deployed to a particular region. It will illustrate tools and techniques that can be used to filter the quality of the collective intelligence, the dynamic categorization of new narrative and the selective recommendation of content as an adaptive learning technique. This system will incorporate a virtual environment to test the quality of learning before the military personnel are deployed and a capture and debrief system that will enable the continual improvement of service personnel as they complete missions during their deployment.

1 Introduction

Military personnel are required to transition to and from assignments in contested cultures many times during their careers of service. This transition usually takes the form of a six-month cycle at the beginning and end of their tour-of-duty. During the transition into an assignment the service person must train on equipment, procedures, communications protocols, and more to become prepared to perform effectively. In addition to their training, they are required to understand the culture they are about to enter in order to effectively communicate, and ultimately survive.

During recent informal interviews with service personnel returning from an assignment overseas, the author also found a significant number of responses indicating that the most effective method for transferring domain and situational knowledge occurred through oral tradition. Individuals reported that team members sharing stories of their

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 611–619, 2009.

experiences with them as they arrived made their transition experience significantly smoother. While some reported that the information conveyed by their predecessor was mistaken, outdated, or misinformed, none reported issues with regards to retaining the lessons conveyed during the transition. In contrast, when queried about current knowledge management systems, service personnel reported many difficulties navigating the often raw and meaningless volumes of content. One service person told the author, “I could either read tons of boring documentation or talk to the guy I was replacing.” It is clear that oral tradition, as a fallback knowledge transition tool is more effective and more popular than current document based knowledge management systems. Unfortunately in its current form it lacks the ability to capture and distribute cumulative knowledge, to provide any record of knowledge transition, and to provide an understanding of the extent of the knowledge has been successfully absorbed. Those relying on oral tradition as an information source are not rewarded by and therefore not concerned about the benefits of capture, distribution, and measurement.

Through the design of a novel system that explores the tenants of social cognitive theory, storytelling, digital systems, and adaptive learning, this paper will attempt to explain not only why oral tradition could be a superior knowledge transition tool, but also how a digital community could be developed that passively capture and adaptively filter community knowledge in a manner that would streamline transitions into a culture and significantly improve situational awareness and continuous learning for individuals currently stationed within contested cultures.

2 Social Cognitive Theory

Social cognitive theory looks to fundamental human behavior as the key to effective learning practices. Human beings, being social creatures, have relied on storytelling techniques for thousands of years as their key to survival. Transferred in the form of “wives’ tales,” tradition, legends, stories, etc., these social cultural norms have served to convey key information in a memorable and applicable manner.

Social cognitive theory began as a body of work that evolved from thinking in the area of social learning proposed by N.E. Miller and J. Dollard in 1941. Miller and Dollard believed that if humans were motivated to learn a particular behavior that particular behavior would be learned through clear observations. By imitating these observed actions the individual observer would solidify that learned action and would be rewarded with positive reinforcement [8].

Social cognitive theory explains psychosocial functioning in terms of triadic reciprocal causation [3]. The theoretical triad illustrates the causal relationship between behavior, cognitive factors, and environmental events as having bidirectional influence on each other. Behavior as an agent is both a stimuli and a response. Traditional Skinnerian behaviorism posits that all behavior is a result of environmental stimuli. Social cognitive theory differs in that it does not claim an originator but instead accepts behavior as a product of other behaviors, environment and cognitive determinism.

Bandura [2] defines social cognitive theory as having three aspects that are particularly relevant to cultural learning and organizational improvement. They are: developing competencies through mastery modeling, strengthening people’s beliefs in their capabilities so they make better use of their talents, and enhancing self-motivation through goal systems.

2.1 Mastery Modeling

Modeling is being widely used with good results to develop intellectual, social and behavioral competencies [3]. The purest definition of modeling is the internalization of concepts through the observation and practice of their demonstrated actions and results. It is the process of internalization that creates strong and memorable associations and hence successful learning.

Systems that utilize exposure to behaviors, environments, and cognitive determinants and then provide reinforcing manners in which to respond and shape personal meaning can provide a development framework that is in alignment with social cognitive theory's mastery modeling. It is, however, important to recognize the difference in observable value between a text-based document and a time-based illustration like a video. Actions illustrated natively, as with a video capture, are easier for the individual to observe and internalize. While observing videos of oneself is common practice, caution should be exercised. Simply being shown replays of one's own behavior usually does not produce much improvement [7]. Observing flawed performances can weaken trainees' beliefs in their capabilities [2].

In addition, successful internalization relies on more than just observation. A successful system must introduce and maintain a social element that provides a framework for individuals to share and practice learning. Social cognitive theory calls this activity "guided skill perfection." At its core, guided skills perfection utilizes social interaction and validation to positively reinforce self-confidence while introducing incremental improvement ideas.

Finally for learning to complete the transition to successful internalization it must be practiced. Much like a second language, level of skill is greatly affected by application and repetition. Social cognitive theory calls this a transfer program. Transfer programs not only provide a "safe" environment to test newly formed skills but also promote advancement by encouraging the undertaking of more and more complex (and difficult) variations of the learned skill as success increases.

3 Storytelling and Oral Tradition

Research showing the positive benefits of storytelling in the field of learning is robust. Cross-culturally, storytelling is a fundamental method of learning even at an early age. Children learn storytelling many years before they master logic, persuasion, writing, and other forms of information delivery. Story is an essential precursor to mastery of expository and logical forms [5]. Humans are, in essence, hard-wired at an early age to transfer learning in the form of stories. Shank reports that storytelling has demonstrable, measurable, positive, and irreplaceable value in teaching [9]. In addition, telling stories is one of the most influential techniques because you give the information, ground the meaning in structure, provide for emotion, and make the content meaningful. Our brain loves storytelling [9]. Narrative details create mental images, making possible both understanding and memory [10]. "Stories enhanced recall, retention, application of concepts into new situations, understanding, learner enthusiasm for the subject matter." and "Stories enhanced and accelerated virtually every measurable aspect of learning" [4].

Storytelling as the presentation form for mastery modeling as part of a social cognitive theory driven approach makes sense and will likely create a vehicle to effectively convey learning, but the question still exists as to how to incorporate guided skill perfection and a transfer program.

4 Designing the Digital Cultural Ecosystem

To begin the design of our novel system, we look to the future of digital community development or specifically the digital cultural ecosystem (DCE). A digital cultural ecosystem is any system that is designed to connect individuals of common purpose and strengthen their shared understanding of key topics, processes, group dynamics, and each other. In short it is a culture that centers on a shared area of interest (SAOI). The SAOI could be meeting people (online dating) or as in the case of our system gathering better understanding about the community in which our troops have been stationed through the eyes of the community that is our troops. DCEs, implemented properly will successfully replace today's knowledge management and intranet systems with relevant, timely, and community validated content feeds.

4.1 Goals for the DCE

As mentioned above, the primary goal of this novel DCE will be to connect our service personnel sharing assignment in contested cultures. It will also:

- Strive to provide participants a better understanding of the culture they inhabit and each other.
- Capture, tag, and filter real-time multimedia information feeds for distribution and storage for further review.
- Be built using adaptive learning principles that work collaboratively with expertise tagging techniques to identify areas of strength for participants and insure they get a well rounded and continuously evolving training program before and during their time on station.
- All opportunities will be taken to minimize the intrusiveness of this system on the daily duties of participants.
- Use emerging technology to measure group interests, activities and areas of focus. Providing administrators the tools to continuously improved the communities cultural dynamic both online and offline.

4.2 Target Culture

The demographic statistics for our target participants will collectively define stationed service personnel:

- Active duty enlisted and officer personnel
- Ages 18 to 40

- Males and Females
- Computer literate with more than 1 year of experience using the Internet.
- Access to a computer terminal or mobile device.

When developing a culture of individuals defined above, care needs to be exercised to insure that participation is not taken for granted. As with most military initiatives there tends to be a "They'll do what their told" mentality when designing systems. While this may or may not be true, the level of quality provided by participation will significantly improved if the participants are excited and rewarded for their behaviors. Part of this reward will need to include entertainment. Much like America's current fixation with reality television, our DCE will use personal accounts and individual personalities to convey cultural learning as perceived.

4.3 Everybody Is a Celebrity

Recent news stories of photo sharing sensitive media by military personnel actually help us understand our target audience a little better in that we can see that they are eager to participate and be validated by their peers in the same way they can on FaceBook or like public community sites. Several publications of recent have begun to discuss the shifting of values among our demographic with regards to privacy and celebrity. In the new world, everyone is a celebrity and shares openly even their most personal details. Systems like Twitter allow others to feel connected to friends or other celebrities as they post glimpses into their current mental state as micro transactions on Twitter.com. The proliferation of digital community systems is a testament to their value as an entertainment, knowledge management, and goals achievement medium. Community systems like FaceBook and LinkedIn have proven valuable tools to connect individuals on personal and professional levels. It's not hard to envision the convergence between emerging technologies and digital community systems.

4.4 Technology Architecture

Instead of text and photos as the primary communication medium our DCE will be built to use stereoscopic video recordings of events and then the narration of the individuals that experienced the events. Additional commentary will be provided by teammates who may further enhance the value of the narrative. The technology stack will include five key components:

1. Helmet or chest armor mounted stereoscopic video recorders (similar to those in today's mountain biker cams). Cameras will be activated during primary activities like patrolling a region or other assignments. Because of the overhead imposed by wireless data transmission, these systems will not transmit video live but instead record and upload wirelessly before debrief sessions.
2. Bio/environmental flagging systems – While the cameras will capture the entire event (up to 16 hours) bio/environmental sensing technology will be used to flag points (time code) on the video that are of interest. Flags will be captured when the wearer feels stress, experiences an elevated heart rate or even is exposed to a loud noise.

3. Debrief stations consisting of:
    a. Wireless downloading facilities
    b. Group presentation and review terminals
    c. Individual and group narrative capture cameras
4. Information distribution and rating systems
    a. Web portal that displays issues geographically and temporally for review by the collective culture.
    b. Measurement systems designed to capture responses from the culture on a specific narrative. This includes direct commentary as well as sharing and blocking activity.
    c. Intelligent tagging architecture that dynamically captures tags from the original narrative (voice-to-text) and the following cultural responses to the narrative.
5. Virtual modeling facility that, by using video capture to derive three-dimensional models, will dynamically build a virtual environment that will allow the replay of events and review from multiple angles.

4.5 A Sample Use Case

Soldier X is about to go on patrol in a contested region. After picking up her gear from the storage locker (uses closed coil technology to charge digital units). She mounts up with her team and begins her assignment. Three hours into the patrol an IED explodes causing no damage to soldier X but injuring one of her teammates. Biosensors tag video as she experiences the explosion and the chaos. Environmental sensors also trigger flags as explosions and shocks are logged as significant events.

As soon as the returning patrol members are within range of the base's wireless network, the camera recording software begins downloading the recorded events. Video footage is compared against teammate footage for correlating events (signifying more importance) and prepared for the debriefing session. When the team is ready for debriefing they enter a debrief circle that consists of a circle of chairs around a display device with screens and cameras for each participant. The debrief unit begins playing back flagged footage and inquires to each team member as to their thoughts and comments about the events that had been flagged. Team members are allowed to add comments to other members' stories further enriching the narrative. Some events are discarded as irrelevant including one when soldier x stubbed her toe while on foot.

When debriefing is complete the remaining aggregated narratives are uploaded to the team portal where they are displayed as new events. The portal displays events as they arrive in real-time in a linear timeline presentation format with the most recent at the top. New events are also geo-tagged with the location of their occurrence and displayed on a map within the portal. Cultural participants download the most recent events and rate those of relevance to their understanding as well as provide feedback, and support for the original narrators. Their responses are captured and used to generate tags that further identify the categorization of the event/narrative. Narratives with more cultural interest receive higher relevance scores using adaptive collaborative filtering (ACF) techniques. Narratives with the greatest volume are pushed to others as items of interest and maintain higher ranking in the collective knowledge pool.

4.6 Adaptive Learning

The use case above references intelligent narrative tagging several times but fails to illustrate the value and purpose of the tagging technology used. To enable intelligent adaptive learning our DCE will capture tags relevant to personal experiences and responses the experiences of others and build a digital "experience" fingerprint of each participant. The fingerprint will infer areas of interest and expertise based on the actions of the participant in the system and the number of tags associated with each topic area. With this fingerprint we can begin to understand areas that need improvement and recommend content that is relevant to those areas as needed. One approach might be to display narratives that are loosely related to the experiences of the participant providing them with insights presented by others having similar experiences. Another might be to look for opposites or deficits in each participant's digital fingerprint and present that material to attempt to fill the gaps. Our design will use both approaches.

4.7 Temporal Dimensional Modeling Repository

Further analysis is often needed for incidents. Video feeds of events and rankings will be converted to a three-dimensional environment using technology very similar to Microsoft's Photosynth technology. Currently Microsoft's Photosynth technology can be connected to still-photo feeds (i.e. Flickr.com) and from them can render dimensional navigable experiences. Figures 1, 2 and 3 illustrate a three-dimensional model of Notre Dame Cathedral derived entirely from an image search done on Flickr.com.

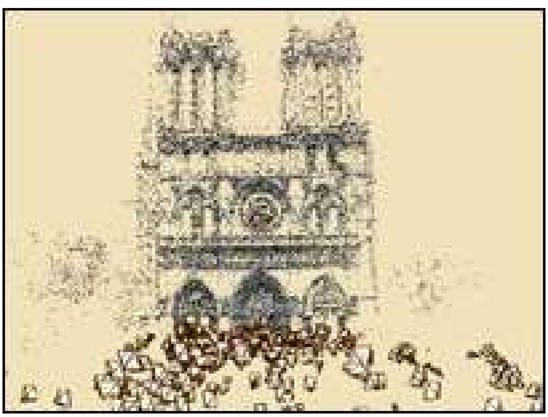


Fig. 1. Photosynth derived Notre Dame mode

This model is fully navigable and presents the viewer with more than a view of the object in a virtual world. It presents a collective memory of the object. Figure 3 illustrates a man standing in front of a poster of Notre Dame that the model accepted as part of its framework and because of it viewers can navigate into the model through the poster as a portal. By grouping photos by date, one can present an accurate history

Fig. 2. Image library used to create Photosynth model

Fig. 3. Notre Dame poster that became an entry point for the Photosynth model

of an object over time. Scaffolding, for instance may only exist for a few weeks and therefore be present in photos dating within those weeks. If one starts to imagine video as thousands of images then the value of the collective memory begins to materialize. Stereoscopic video presents us the ability to create images of individuals that can be navigated around. Viewers in our collective memory will be able to walk around slow moving or still images of individuals as they experience events creating opportunities for analysis and even training/teaching/learning.

5 Summary

The possibilities are exciting when one starts to think of the possibilities presented by a real-time capture and modeling system that uses the eyes and experiences of every single participant as input. Rich and robust knowledge and even wisdom can be generated by just one modeling session. It is important to remember that all digital systems must provide insights into the activity they host. These insights are as valuable, if not more so, than the original data that is captured.

References

1. Aguera y Arcas, B.: Photosynth Demo at Ted Talks (2007), http://www.ted.com/index.php/talks/blaise_aguera_y_arcas_demos_photosynth.html
2. Bandura, A.: Organizational Application of Social Cognitive Theory. Australian Journal of Management (1988)
3. Bandura, A.: Social Foundations of Thought and Action: A Social Cognitive Theory. Prentice-Hall, Englewood Cliffs (1986)
4. Coles, R.: The Call of Stories. Houghton Mifflin Company, Boston (1989)
5. Engle, S.: The stories children tell: Making sense of the narratives of childhood. Freeman, New York (1995)
6. Gross, R., Acquisti, A.: Information revelation and privacy in online social network (the facebook case). In: Proceedings of the Workshop on Privacy in the Electronic Society, Nottingham, UK (September 2004)
7. Hung, J.H., Rosenthal, T.L.: Therapeutic Videotaped Playback. In: Fryrear, J.L., Fleshman, R. (eds.) Videotherapy in Mental Health. Thomas, Springfield (1981)
8. Miller, N.E., Dollard, J.: Social Learning and Imitation. Yale University Press, New Haven (1941)
9. Shank, R.C.: Tell me a story: A new look at real and artificial memory. Scribners, New York (1990)
10. Tannen, D.: The argument culture: Moving from debate to dialogue. Ballantine, New York (1999)

The Interaction between Chinese University Students' Computer Use and Their Attitudes toward Computer in Learning and Innovation

Ye Liu and Xiaolan Fu*

State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China
{liuye,fuxl}@psych.ac.cn

Abstract. A survey study investigated the dynamic interaction between Chinese university students' computer use and their attitudes toward computer in learning and innovation. The relationships among attitudes toward computer in learning (ACL), attitudes toward innovation (ATI), and self-perception on computer skill (SPCS), were also examined. Participants were 292 university students from three universities in Beijing, and all of them had used computer and Internet before. The results showed that: (a) Previous computer use could predict recent computer use with ACL, ATI, and SPCS as mediate variables; (b) males were more confident than females in SPCS, but there was no gender difference in either ACL or ATI; and (c) the participants' notion of innovation was significantly more positive than their innovative action.

Keywords: Computer use; Computer attitude; Attitudes toward innovation; Self-perception on computer skill; Gender difference.

1 Introduction

The 14th Statistical Survey Reports on the Internet Development in China (China Internet Network Information Center, July, 2008) showed that 30.3% of the Internet users' were between the ages of 18 and 24 in China, and 30.0% of the Internet users were students. These data implicated that university students were the biggest group of Internet users in China. In fact, university students were also potentially the biggest group of computer users. Undoubtedly, Computer and Internet have made great influence on Chinese university students' learning during the last decade. However, the relationship among Chinese university students' computer use and the related factors has not yet received enough attention from researchers.

There have been a lot of studies all over the world, especially in Europe and America, which addressed the issues of computer use and the related factors, such as attitudes toward computer, attitudes toward innovation, and self-efficacy [6]. However, it is not easy to know how their results and conclusions can be generalized to Chinese,

* Corresponding author.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 620–629, 2009.

because the culture of China is so different from those of other countries. It is necessary to investigate Chinese university students' computer use and the related factors. Recently, computer and internet are more and more popular and widespread in China. The behavior and attitudes toward technology of Chinese are enormously influenced by the use experience. In the present study, the main studies on computer use, attitudes toward computer and innovation are reviewed firstly. And then, this paper reports findings from a questionnaire survey designed to investigate the interaction between Chinese university students' computer use and their attitudes toward computer in learning and innovation.

1.1 Computer Use

Smith, Caputi, Crittenden, Jayasuriya, and Rawstorne [18] have reviewed and evaluated the construct of computer experience. They differentiated objective computer experience and subjective computer experience and proposed a bi-dimensional view of computer experience. The following study [19] provided empirical proof for the bi-dimensional view. Since objective and subjective experience are different concepts, computer use as a measure of objective computer experience was investigated in the present research. According to Smith et al. [18], Computer use in the present study was defined as the totality of externally observable, direct human-computer interactions that transpire across time.

Some researchers consistently found that computer use influenced on computer attitude or had positive relationship with computer attitude [1, 3, 6, 9, 10, 13, 15, 19]. However, most of the previous studies measured computer use by scores of Likert scale, and they didn't differentiate previous computer use and recent computer use. In the present study, previous computer use and recent computer use was differentiated to investigate the effect of previous experience on recent experience. Previous computer use was measured by estimating computer use history, and recent computer use was measured by estimating the time of computer using.

1.2 Attitudes toward Computer

Most theorists agreed that evaluation constitutes a central aspect of attitudes. Affective, cognitive, and behavioral antecedents of attitudes can be distinguished, as can affective, cognitive, and behavioral consequences of attitudes [14]. The affective component of attitude is the emotion or feeling which includes statements of likes or dislikes about some certain objects. The cognitive component is statements of beliefs that a certain object can increase significantly the quality of her/his output. And the behavioral component is what an individual actually does or intends to do [10]. According to this idea, the Computer Attitude Measure (CAM), developed by Kay [7], was composed of demographic information, cognitive, affective, and behavioral attitudes. The Computer Attitude Scale, developed by Loyd and Loyd [11], consisted of computer anxiety, computer confidence, computer liking, and computer usefulness.

These studies measured attitudes toward computer from three components, but the central aspect of attitudes, evaluation, was neglected. Eagly and Chaiken (1992) claimed attitudes do not form until individuals respond evaluatively to an entity (see 14). Based on this opinion, attitudes toward computer in learning (ACL) were consisted of evaluation, cognitive, affective, and behavioral component in the present study.

Whitely [20] found that males specifically rated themselves higher than females on the following two components: Computer competence and gender-appropriateness (see 13). Therefore, self-efficacy was separated from computer attitude and was measured as self-perception on computer skill (SPCS) in the present study.

1.3 Attitudes toward Innovation

The relationship among attitudes toward innovation, computer attitude, and computer use has attracted some researchers' attention [2, 4, 5, 12]. Adaption-innovation theory [8] assumed a continuum of cognitive styles on which individuals were positioned according to their preference for a manner of decision-making, problem-solving, and creativity that emphasized relative continuity and incremental change (the adaptive pole) or relative discontinuity and radical change (the innovative pole) (see 5). Researchers (e.g., 4; 5) found that the number of applications to which the computer was put and the frequency of computer using correlated with total scores on the Kirton Adaption-Innovation Inventory. Another innovativeness theory described innovativeness as a personality characteristic that indicated how relatively early an individual was in adopting an innovation relative to others in a social system (see Braak, 2001).

Attitudes toward innovation (ATI) were defined based on Braak's research [2] in the present study. Different from the previous studies, innovativeness was measured from two dimensions in the present study: Notion of innovation and action of innovation. In Chinese culture, golden mean is canonized. Although foreign cultures had great influence on Chinese students' notion, their behavior should be a little conservative. Therefore, the hypothesis was that Chinese students' notion of innovation may be at variance with their action of innovation.

1.4 Research Questions in Present Study

Based on the previous research reviewed above, the specific research question in the present study was addressed as follow:

(1) What kind of the distribution of hours on computer using of Chinese university students differed in gender, grade, and major?
(2) Which of the three variables, gender, grade, and major, were possible factors to influence attitudes toward computer and innovation, and self-perception on computer skills?
(3) What kind of possible relationship was among these three aspects and computer use?

In the present study, previous computer use (years of computer using) and recent computer use (hours per day on computer recently) were differentiated. The hypotheses were that previous computer use may influence ACL, ATI, and SPCS, and then the three aspects may influence recent computer use in turn.

2 Method

2.1 Participants

The questionnaires were distributed to three hundred university students in random at the main libraries and classrooms of three universities in Beijing. The percentage of valid participants was 97.3%. In these 292 students (162 males and 130 females), 100 were from Peking University, 99 from Tsinghua University, and 93 from Beijing Normal University. Their ages ranged from 16 to 25 years (M = 20.28). All participants had experience of using computer.

2.2 Instruments

The survey questionnaire consisted of investigation of basic information and three sub-scales.

Basic information. Basic Information consisted of gender, age, major, grade, years of computer using, hours on computer per day recently.

Three sub-scales. Computer Attitude Scale, Self-Perception Scale, and Innovativeness Scale were used. All of the 106 items used 6-point Likert scale (strongly disagree, much disagree, little disagree, little agree, much agree, and strongly agree). All of odd number items were positive statements, and even number items were negative statements. The data of the negative statements were reversed before statistical disposing. The split-half reliability of the whole scale was 0.86.

a) Computer Attitude Scale. It measured ACL, and included four dimensions: cognitive component (22 items, $\alpha = .90$), behavioral component (22 items, $\alpha = .87$), affective component (16 items, $\alpha = .87$), and evaluation (22 items, $\alpha = .90$).
b) Self-perception Scale. It measured SPCS, and consisted of 4 items ($\alpha = .81$).
c) Innovativeness Scale. It measured ATI from two different dimensions: Notion (11 items, $\alpha = .69$) and action (9 items, $\alpha = .70$).

3 Results

3.1 Computer Use

Most of participants had been using computer for a few years (M = 5.28 years, SD = 2.66). A 2×2×3 analysis of variance (ANOVA) was performed on the years of computer using with gender, grade, and major as factors. The results showed that females had used computer significantly longer than males, 5.60 vs. 4.90, F (1, 279) = 4.47, $p < .05$. The students in high grades had used computer longer than the students in low grades, 5.55 vs. 4.95, F (1, 279) = 3.26, $p = .07$. The main effect of major was also significant, F (2, 279) = 5.02, $p < .01$. The post hoc test revealed that the students who majored in Natural Science had used computer significantly longer than the students either in Humanity and Social Science or in Engineering (6.05 vs. 4.80, 4.90, respectively). All the interactions were not significant, $Fs < 2.00$, $ps > .05$.

The average time of using computer per day was 1.61 hours (SD = 1.37). The results showed that the main effect of gender was not significant, $F(1, 272) = 2.87$, $p = .09$. The main effect of grade was significant, $F(1, 272) = 22.38$, $p < .001$. The students in high grades spent significantly more hours on computer than the students in low grades (1.84 vs. 1.03). The main effect of major was significant, $F(2, 272) = 6.30$, $p < .01$. The post hoc test revealed that the students majoring in Engineering spent significantly more hours on computer per day than those either in Humanity and Social Science or in Natural Science (1.89 vs. 1.34, 1.39, respectively). All the interactions were not significant, $Fs < 1.70$, $ps > .05$.

3.2 Three Sub-scales

The data showed Chinese university students' attitudes toward computer in learning were slightly positive (M = 4.36, SD = 0.64), their attitudes toward innovation were apt to positive (M = 3.99, SD = 0.59), and their self-perceptions of computer skill were slightly negative (M = 2.98, SD = 1.16).

A 2×2×3 multivariate analysis of variance (MANOVA) was conducted to determine the effects of gender, grade, and major on the average of the three sub-scales: ACL, ATI, SPCS. The differences between gender were significant, the value of Pillai's Trace is 0.04, $F(3, 278) = 4.08$, $p < .01$. Significant differences were found between grades on the dependent measures, the value of Pillai's Trace is 0.11, $F(3, 278) = 11.86$, $p < .001$. The differences among majors were also significant, the value of Pillai's Trace is 0.06, $F(6, 558) = 2.93$, $p < .01$. Analysis of variance (ANOVA) on each dependent variable was conducted as follow-up tests to the MANOVA. The difference between gender was significant on self-perception on computer skill, $F(1, 280) = 10.22$, $p < .01$. Males were more confident than females (3.15 vs. 2.80). The difference between grades was significant on self-perception on computer skill, $F(1, 280) = 35.82$, $p < .001$. The students in high grades were more confident than the students in low grades (3.33 vs. 2.51). The differences among majors were also significant on self-perception on computer skill, $F(2, 280) = 4.81$, $p < .01$. The post hoc test revealed that students majoring in Natural Science were more confident than students majoring in Humanity and Social Science (3.33 vs. 2.81). The differences among majors were also significant on innovativeness, $F(1, 280) = 3.45$, $p < .05$. The students majoring in Engineering and in Natural Science were more innovative than those in Humanity and Social Science (4.02, 4.05 vs. 3.82, respectively). All the other main effects were not significant, $Fs < 3.35$, $ps > .05$. All the interactions were not significant, $Fs < 2.10$, $ps > .05$.

The comparison between notion and action of Innovation was reliable, $t(291) = 23.36$, $p < .001$. The scores of notion were higher than scores of action (4.54 vs. 3.60).

3.3 Path Analysis to Explain the Relationship among Computer Use and the Three Sub-scales

Path analysis was used to examine the relation among computer use, ACL, ATI, and SPCS. The basic model and results of the path analysis are presented in Figure 1.

As indicated in Figure 1, previous computer use influenced SPCS, and then SPCS and ACL directly influenced recent computer use. Although ACL and ATI had no

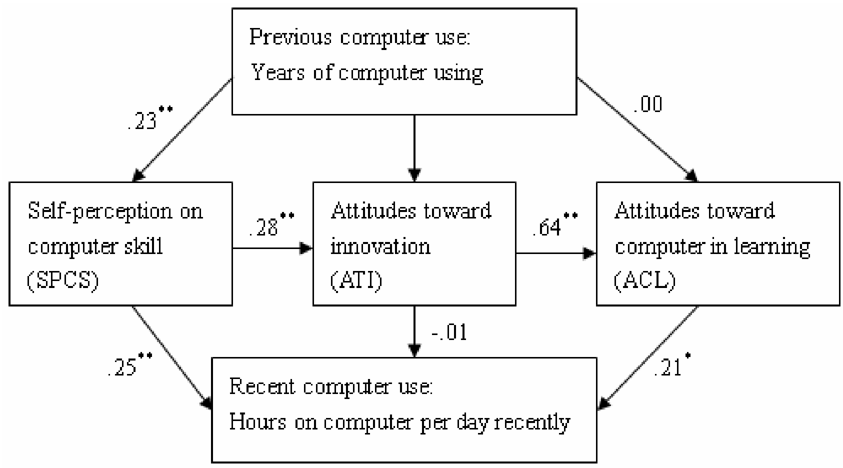


Fig. 1. Path-analytic model: The relationships among computer use, ACL, ATI, and SPCS. Paths are standardized coefficients. $*p < .05$, $**p < .01$.

significant relationship with previous computer use, previous computer use had indirect influence on ACL and ATI by SPCS. Previous computer use could predict recent computer use with ACL, ATI, and SPCS as mediate variables. Six indices used to evaluate the hypothetical model supported this conclusion, $\chi^2(2) = 1.293$, $p = .524$; root mean square error of approximation (RMSEA) = 0.000; comparative fit index (CFI) = 1.000; Tucker-Lewis index (TLI) = 1.001; Incremental fit index (IFI) = 1.000; Normed fit index (NFI) = 1.000. Except the paths from previous computer use to ACL and ATI, and the path from ATI to recent computer use, the standardized parameters of other paths in the model were statistically significant ($p < .05$).

4 Discussion

The present study found that previous computer use could predict recent computer use with ACL, ATI, and SPCS as mediate variables. Males were more confident than females in SPCS, but there was no difference between males and females on either ACL or ATI. These university students' notion of innovation was more radical than their action.

4.1 Computer Use

The present study found that the university students in this survey general spent 1.61 hours per day on computer. In contrast, Male students in American university spent 1.32 hours on computer per day, and female students in American university spent 0.86 hours on computer per day as reported in the study by Schumacher and Morahan-Martin [16]. Comparatively, the Chinese university students as measured by this survey had much more experience on computer. The explanation for this finding can perhaps be attributed to the rapid popularization of computers in China recently.

Females had used computer significantly longer than males (5.64 vs. 4.98 years). But no gender difference was found on hours spent on computer and Internet. The differences among majors were significant. The students majored in Natural Science had used computer significantly more years than the students either in Humanity and Social Science or in Engineering. But the students majoring in Engineering spent significantly more hours per day on computer and Internet than those either in Humanity and Social Science or in Natural Science. Comparison across grades showed that the students in higher grades spent more hours than those in lower grades. It seems that the students may use computer more and more along with the difficulty and profundity of courses increases.

The results of path model indicated previous computer use influenced SPCS, and then SPCS directly influenced recent computer use. Although ACL and ATI had no significant relationship with previous computer use, previous computer use had indirect influence on ACL and ATI by SPCS. ATI influenced recent computer use through ACL. Overall, previous computer use could predict recent computer use with ACL, ATI, and SPCS as mediate variables.

Such results are consistent with the findings from a lot of previous research (e.g. 3; 9), which showed that computer use had positive relationship with computer attitude, self-efficacy, and attitudes toward innovation. In the present study, probably because the Chinese university students had used computer for a long time, and the three universities were well equipped advanced computers and Internet hardware, their computer experience made them feel confident in computer technology. Their self-efficacy and confidence would improve the level of their attitudes toward innovation from both of notion and action. If they were willing to accept new computer technology, then they should realize the importance of mastering technology. Once they had fairly positive attitudes toward computer so that they were willing to use computer. According to attitude-behavior theory (see 9), attitude is an important precursor of behavior. The positive attitudes toward computer will arise more positive behavior on computer using in learning. Therefore, colleges and universities should pay more attention to computer education in order to shape the positive attitudes of students toward computer and further facilitate their behavior on computer using in learning.

4.2 Attitudes toward Computer in Learning

Overall, the results of present study indicated that the participants were willing to take positive attitudes toward computer in learning. There was no difference between males and females on attitudes toward computer in learning. Both of males and females had much experience on computer and recognized that technology was very serviceable to them. Consequently, they were willing to use technology and felt fairly comfortable and satisfied with using technology. It strongly suggests that females and males would have equal opportunity to enjoy the convenience from hi-tech equipment.

4.3 Self-perception on Computer Skill

The participants' self-perceptions of computer skill were a little bit negative, though they had used computer for a few years. But this result does not necessarily mean that

Chinese university students were not confidence in their computer skill. Such result might have something to do with the Chinese culture. In China, humility is usually an expected virtue for everyone, especially in the education of the youngsters. In Chinese traditional culture, youngsters are required to be modest, humble, and open. So, the participants in the present research didn't give higher rating to the computer skill of themselves.

It is worth noting that, although females had used computer previously for a longer time, and they felt as same comfortable and satisfied with using technology as males, but they were short of confidence in their computer skill. The results are consistent with Whitely's [20] finding. Considering the custom that females are (or culturally perceived to be) more modest than males in eastern Asia, we may understand that there was no gender difference in attitudes toward computer, while there was gender difference in self-perception on computer skill. In Chinese traditional culture, the virtue of females that people praised is goodness, peaceful, humility, and forbearing. Based on the tradition, the female participants in the present research would be more modest than the male participants, and give even lower rates in their computer skill.

The results also showed major difference and grade difference in self-perception on computer skill. Students in Natural Science were more confident than those in Humanity and Social Science. It is not surprised because the students in Natural Science usually have more opportunities to use computer and Internet. Seniors were more confident in their computer skill. It is true that college time is an important stage, specifically as access to hi-tech learning method, for most of Chinese students. As their experience and knowledge enriched gradually, the students may improve their skills in computer and become more confident. However, the details of the explanation need further investigation.

4.4 Attitudes toward Innovation

The results showed that participants' attitudes toward innovation were inclined to positive, and had a significantly effect on ACL. It is consistent with Braak's [2] findings and suggests that attitudes toward innovation may be a strong predictor of ACL.

We found, most notably, the notion of innovation of these university students was more radical than their action. Such pattern may emerge from the fact that the students are really willing to adopt all kinds of new hi-tech products, but they have no much chance to access them. Although this hypothesis still need testing, it is highly possible that more opportunities supplied to the university students would have positive influence on making their potential of actions become true. On the other side, such results might indicate that the university students' attitudes to innovation could be affected by the radical innovative action of their peers; therefore reliable instruction should be given to the students and lead them to reasonably use new technology.

The present study showed computer use experience had significant influence on Chinese undergraduates' attitudes toward technology and innovation. The implications of the result may lie in twofold. First, Chinese students are inclined to accept new technology, and this is very positive to the promote computer use in China. Second, there should be some reliable restricts to avoid their using computer and Internet too much. Nowadays, the addiction to Internet becomes a serious problem in adolescents, and it is another important social issue.

Acknowledgements

This research was supported in part by grants from 973 Program (Grant No. 2006CB303101), the National Natural Science Foundation of China (Grant No. 30700233), and the Project for Young Scientists Fund, Institute of Psychology, Chinese Academy of Sciences (07CX132013). The authors wish to thank Dr. Winton AU Wing Tung of Chinese University of Hong Kong for his valuable comments and suggestions on the early draft of this manuscript.

References

1. Bozionelos, N.: Computer anxiety: relationship with computer experience and prevalence. Computers in Human Behavior 17(2), 213–224 (2001)
2. Braak, V.J.: Individual characteristics influencing teachers' class use of computers. Journal of Educational Computing Research 25(2), 141–157 (2001)
3. Cassidy, S., Eachus, P.: Developing the Computer User Self-Efficacy (CUSE) Scale: Investigating the relationship between computer self-efficacy, gender and experience with computers. Journal of Educational Computing Research 26(2), 133–153 (2002)
4. Foxall, G.R., Bhate, S.: Psychology of computer use: XIX. Extent of computer use: Relationships with adaptive-innovative cognitive style and personal involvement in computing. Perceptual and Motor Skills 72(1), 195–202 (1991)
5. Foxall, G.R., Hackett, P.M.: Cognitive style and extent of computer use in organizations: Relevance of sufficiency of originality, efficiency and rule-conformity. Perceptual and Motor Skills 74(2), 491–497 (1992)
6. Garland, K.J., Noyes, J.M.: Computer experience: a poor predictor of computer attitudes. Computers in Human Behavior 20(6), 823–840 (2004)
7. Kay, R.H.: An exploration of theoretical and practical foundations for assessing attitudes toward computers: The computer attitude measure (CAM). Computers in Human Behavior 9(4), 371–386 (1993)
8. Kirton, M.J.: Adaptors and innovators: a description and measure. Journal of Applied Psychology 61, 622–629 (1976)
9. Levine, T., Donitsa-Schmidt, S.: Commitment to learning: Effects of computer experience, confidence and attitudes. Journal of Educational Computing Research 16(1), 83–105 (1997)
10. Liaw, S.S.: An Internet survey for perceptions of computers and the World Wide Web: relationship, prediction, and difference. Computers in Human Behavior 18(1), 17–35 (2002)
11. Loyd, B.H., Loyd, D.E.: The reliability and validity of instruments for the assessment of computer attitudes. Educational and Psychological Measurement 45, 903–908 (1985)
12. Mitra, A., LaFrance, B., McCullough, S.: Differences in attitudes between women and men toward computerization. Journal of Educational Computing Research 25(3), 227–244 (2001)
13. North, A.S., Noyes, J.M.: Gender influences on children's computer attitudes and cognition. Computers in Human Behavior 18(2), 135–150 (2002)
14. Olson, J.M., Zanna, M.P.: Attitudes and attitude change. Annual Review of Psychology 44, 117–154 (1993)
15. Rozell, E.J., Gardner, W.L.: Cognitive, motivation, and affective processes associated with computer-related performance: a path analysis. Computers in Human Behavior 16(2), 95–110 (2000)

16. Schumacher, P., Morahan-Martin, J.: Gender, Internet and computer attitudes and experience. Computers in Human Behavior 17(1), 95–110 (2001)
17. Semiannual Survey Report on the Development of China's Internet recommended by China Internet Network Information Center (July 2008), `http://www.cnnic.net.cn/download/2008/CNNIC22threport-en.pdf` (retrieved August 8, 2008)
18. Smith, B., Caputi, P., Crittenden, N., Jayasuriya, R., Rawstorne, P.: A review of the construct of computer experience. Computers in Human Behavior 15(2), 227–242 (1999)
19. Smith, B., Caputi, P., Rawstorne, P.: Differentiating computer experience and attitudes toward computers: an empirical investigation. Computers in Human Behavior 16(1), 59–81 (2000)
20. Whitely, B.: Gender differences in computer related attitudes and behavior: A meta-analysis. Computers in Human Behavior 13(1), 1–22 (1997)

Peak Performance Trainer (PPT^{TM}): Interactive Neuro-educational Technology to Increase the Pace and Efficiency of Rifle Marksmanship Training

Giby Raphael[1], Chris Berka[1], Djordje Popovic[1], Gregory K.W.K. Chung[2], Sam O. Nagashima[2], Adrienne Behneman[1], Gene Davis[1], and Robin Johnson[1]

[1] Advanced Brain Monitoring, 2237 Faraday Ave., Suite 100, Carlsbad, CA 92008
chris@b-alert.com
[2] UCLA/National Center for Research on Evaluation, Standards, and Student Testing: CRESST

Abstract. Marksmanship training involves a combination of classroom instructional learning and field practice involving the instantiation of a well-defined set of sensory, motor and cognitive skills. Current training procedures rely heavily on conventional classroom instruction often with qualitative assessment based on observation (i.e. coaching). We have developed a novel device called the Peak Performance Trainer (PPT^{TM}) which can accelerate the progression from novice-to-expert based on automated inferences from neurophysiological measurements. Our previous work has revealed specific EEG correlates to stages of skill acquisition in simple learning and memory tasks. We have incorporated this knowledge as well as an array of other physiological metrics to develop a field-deployable training technology with continuous physiological monitoring in combination with simultaneous measures of performance, workload, engagement and distraction, accuracy, speed and efficiency. This paper outlines the features of the PPT and the preliminary results of its use in marksmanship training.

Keywords: EEG, Heart rate, Alpha, Theta, Haptics.

1 Introduction

Skill development is thought to occur in stages characterized by distinctive amounts of time and mental effort required to exercise the skill: the initial cognitive stage of assembling new knowledge, the associative stage where newly assembled procedural steps gradually automate as they are practiced, and the autonomous stage where the task execution is automated and performed with minimal conscious mental effort [1-3]. During the transition from the cognitive to associative stage, both speed and accuracy increase as subjects become less reliant on the declarative representations of knowledge [4-5]. Transitions between stages can be assessed with expert observations and subjective reports but these measures often lack precision and do not offer insight into the neurocognitive processes involved during learning. Recent investigations suggest that changes in EEG power spectra and event-related EEG can be identified as associated with stages of skill acquisition in simple and complex tasks [6-8].

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 630–639, 2009.

Relationships between EEG parameters and proficiency in real world activities have been reported in golf putting [9], archery [10], and marksmanship [11-13]. In these real-world task environments, the most predictive data is acquired during the period of mental preparation (usually between 8-15 seconds in duration) before the skilled movements occur, referred to in sports medicine as the "pre-shot routine" [9]. The pre-shot routine is characterized by a progressive increase of the power of EEG in the alpha bands (8-12Hz) particularly over the parietal-occipital regions, with decreased activation in cortical regions not relevant to skilled visuomotor tasks [14,15]. Alpha power in expert marksmen is particularly increased over the left central-temporal-parietal region during the seconds preceding trigger pull [14,16]. The magnitude of the increase in pre-shot alpha power has been positively correlated with the accuracy of the subsequent shot [15,17] n both experts and novices. Less EEG activation is observed across all brain regions for experts compared to novices, suggesting that the neural networks of experts may be more efficiently organized than novices providing a relative economy in the recruitment of cortical resources in the expert brain [14]. The pre-shot period is also characterized by heart rate deceleration and a decrease in electrodermal skin conductance levels [18-20]. Heart rate changes are believed to reflect the focusing of attention and the skill-related aspects of sensory-motor preparation for performance [21]. Consistency and reproducibility of the successful pre-shot routine is a major feature that distinguished novice from expert [22, 23].

Our previous work has revealed specific EEG correlates of stages of skill acquisition in simple learning and memory tasks and in more cognitively complex and challenging test environments. Unique event-related EEG signatures detected during various stages of skill acquisition were evaluated to assess participants' ability to reflect aspects of learning across tasks and environments. The EEG-engagement measure has been shown to correlate with the number and complexity of visual stimuli being processed and the allocation of attentional resources in simulation tasks [24-26]. EEG-engagement increased as a function of level of interest in a specific display (equally sensitive and specific for text or image-based presentations) as well as during the encoding period of memory tasks and during review of instructions for completing a new task. EEG-engagement and workload levels decreased as a function of increasing level of skill acquisition [26, 27].

Transition from novice to expert requires practice. Repetition alone however, does not ensure success and a poor technique repeated can lead to performance deficiencies and/or stress injuries. Instructional strategies and feedback are believed to be critical to accelerating motor skill learning. Recent investigations have suggested that motor skill learning may be dependent upon the availability of cognitive resources including attention and working memory and that the speed and efficiency of learning may be affected by either state or trait differences in these cognitive capacities [28, 29].

The Peak Performance Trainer (PPT) is a novel system that incorporates our knowledge about EEG signatures as well as an array of other physiological metrics that change during stages of learning. The goal of the PPT is to provide continuous psychophysiological monitoring and feedback (visual, auditory or haptic) on relevant changes in these measures to the trainee in real time. Our hypothesis is that we can characterize the psychophysiological profile of expertise, and provide feedback to shape the novice into the psychophysiological state of an expert. The laboratory-based

PPT was designed to offer multiple options for training including: sensor inputs (EEG, EKG, respiration, eye tracking), algorithms for deriving state changes (based on single or multiple sensor inputs, designed for shaping to an expert model) and feedback delivery (visual, audio, haptic or multimodality). Training can be customized to meet the needs of the investigators or the trainees. The training protocols can then be streamlined and optimized for field deployability. The mobile PPT is then designed for portability with fewer options for sensors and/or feedback.

2 Methods

In order to first assess the concept and effectiveness of using the PPT to accelerate learning, a PPT lab setup was designed, built, and pilot tested on a group of novice participants in a current marksmanship study.

2.1 PPT Apparatus

Shooting was untimed, completed indoors in the kneeling position, and simulated a 20" target at 200 yards distance. A demilitarized "airsoft" replica of the M4 was used as the instrumented weapon and an infrared laser-based training system from LASERSHOT [30] was used for target projection (via an LCD projector) and shot detection (via infrared camera). Psychophsysiological metrics associated with shooter performance were used to compare experts to novices, to examine the efficacy of interventions based on these metrics and to guide interventions leading to rapid skill acquisition. An overview of the recorded measures, with their respective source and usage is listed in Table-1.

All necessary sensors and associated electronics incorporated in the PPT setup were integrated into a portable package that provided both closed loop real-time feedback to the shooter as well as transmitted data to a remote computer for display, storage and offline analysis. A low power 32-bit ARM9 processor was used to interface the sensors, for software signal processing and for running the complex feedback algorithms. The sensors and data acquisition circuitry of a previously developed 9-channel wireless B-Alert® sensor headset were integrated to the microprocessor to acquire high quality physiological signals such as EEG from sensors placed at F3, F4, C3, C4, P3, P4, Fz, Cz and POz positions (according to the international 10-20 system) as well as EKG and EOG. The patented EEG sensor dispenses a small amount of conductive cream through the hair to make electrical contact, which eliminates the need for hair or scalp preparation. The sensors and the headset were attached to a comfortable porous cap. Analog circuits combined with EEG amplification close to the sensors allowed the shooter freedom to move without generating artifacts.

The acquired data was processed to identify and decontaminate artifacts using hardware filters as well as adaptive filters in the firmware. The filtered data was then used by the algorithms running in the microprocessor for real-time feedback and also transmitted to a remote computer via Bluetooth protocol for display, storage and analysis. The Bluetooth module was interfaced to the microprocessor via the serial port to transmit the data, and an off-the-shelf Bluetooth dongle plugged in to the USB

Table 1. Metrics recorded for analysis

Metric	Data Source	Usage
Cognitive overload Pre-Shot EEG Alpha and Theta	EEG	Used as an indicator of how well shooter is processing information and accommodating task demands. Used as an indicator of focused and relaxed ("in the zone") mental state
Anxiety	Heart rate variability	Used to measure degree of stress experienced by shooter.
Precision	Shots	Used to characterize the degree of dispersion of shots.
Accuracy	Shots	Used to characterize the distance of shots from the intended target.
Respiration	Breathing	Used to measure inhalations and exhalations.
Trigger break	Switch	Used to establish a synchronization point for all measures.
Trigger squeeze	Force pressure sensor	Used to examine quality of trigger squeeze - slow or rapid.
Muzzle wobble	Accelerometer	Used to measure the degree of movement in the muzzle of the weapon.

port of the computer was used to receive the data at the remote end. Our patented B-Alert software was used for display and analysis of the data on the computer. The B-Alert software provided capabilities for further processing of the data such as classifying the brain's electrical activity into validated measures of engagement, mental workload, and distraction/drowsiness. The package had multiple analog and digital input ports to interface external sensors. Three primary real-time feedback modalities were provided: i) audio feedback via a small speaker interfaced to the microcontroller as well as from the remote computer; ii) haptics feedback via two shaft-less vibration motors attached behind the neck; and iii) visual display on a projected screen and the remote computer.

Alpha power in expert marksmen has been shown to increase during the seconds preceding trigger pull [14, 15]. Preliminary studies suggested that a single EEG channel is sufficient for the measurement of alpha increase, and that alpha (as well as theta) levels in the two seconds preceding the shot were at least 10% higher than during inter-shot interval in experts. In order for the PPT to quantify pre-shot theta and alpha levels and provide that feedback to the shooter, band pass filters were used to extract theta (4-7Hz) and alpha (8-12Hz) bands from channels Fz and Cz of the EEG signal and the area under the curve (squared) was calculated to extract the energy per unit time of the bands. Individualized baseline alpha levels were calculated during a 30-second interval before shooting began, and real-time feedback of the current alpha levels during the shooting process was provided using audio as well as haptic vibrators. The alpha levels were classified as good (less that 5% above baseline), better (between 5% and 10% above baseline) and best (10% above baseline). Haptic feedback was provided using the potential feedback states: two vibrators were both switched ON when alpha levels were at or below normal; one stopped vibrating when alpha level increased to 5% above normal; and both stopped vibrating when alpha increased the at least 10% above normal – indicating to the participant that they were in an ideal state and ready to shoot the weapon. Some initial training was required for subjects to learn to increase their alpha levels and to rapidly move in and out of the desired pre-shot brain state.

Significant deceleration of HR was found in experts beginning about two seconds pre-shot and ending at around 0.5 seconds following shots. Heart rate (HR) and Heart Rate Variability (HRV) measures add the dimensions of stress, frustration, arousal and anxiety to the EEG-based cognitive metrics. In order for the PPT to provide HR feedback to the participant, missed beats and false detections due to movement artifacts in the EKG were corrected using software filters. A proprietary adaptive R-wave detection algorithm was employed to detect the R-waves in the EKG to provide real-time feedback of the heartbeat to the shooter. The haptic vibration pattern for the alpha feedback was triggered by R-wave detection (the haptic motors vibrated at an interval that was in synch with the heart beat) thus superimposing the alpha feedback over the heartbeat feedback. This reduced the need to memorize complex feedback patterns for the shooter.

Breath control is another significant factor in marksmanship; firing during the natural respiratory pause is accepted as the correct procedure. The expansion and contraction of the lungs during the breathing cycle can cause the rounds to be dispersed vertically on the target due to the displacement of the muzzle. A respiratory belt transducer containing a piezo-electric device that responded linearly to changes in length was used to measure changes in thoracic or abdominal circumference during respiration. The necessary circuitry for the amplification and filtering of the signal was implemented in hardware and the transducer was interfaced to the analog input of the microprocessor. The signal representing the respiratory pattern was transmitted to the remote computer and was stored and displayed in real-time.

Trigger control was measured by determining the force profile of the shooter's trigger finger during the aiming period immediately preceding the shot (about 6 seconds) and immediately following the shot. Proper trigger control is important because yanking the trigger will cause the weapon to sway laterally. A force pressure sensor was attached to the trigger and the resultant pressure on the sensor measured over time. The signal was amplified, filtered and transmitted to the remote computer for real-time display and storage. The respiratory and the trigger squeeze waveforms were superimposed to detect variations from the expert trigger squeeze profile which consists of a steady increase in trigger pressure during natural exhalation with the trigger break occurring at the end of the exhalation and before the beginning of the next inhalation (Fig. 1). We also provided audio feedback of the trigger squeeze profile by modulating the frequency of the sound with the amplitude of the pressure applied on the trigger.

Muzzle wobble can never be completely eliminated, however its magnitude was found to be considerably depressed in experts. An accelerometer was attached to the muzzle of the instrumented weapon to measure the degree of movement in the muzzle of the weapon. The signal from the accelerometer was amplified and transmitted to the remote computer for real-time display.

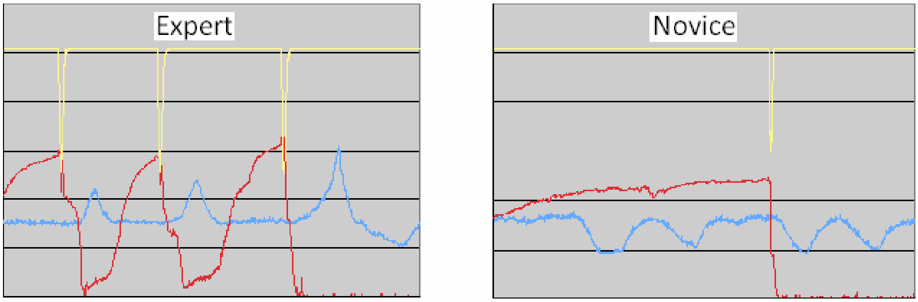


Fig. 1. Snapshots of the real time display from expert and novice shooters, showing trigger break (*yellow*), trigger squeeze profile (*red*), and breathing profile (*blue*)

Shots were recording using a LaserShot rifle simulation set-up. The target was projected on a screen and the laser "shot" was detected by a high-resolution infrared camera. The strike was interpreted by the system to yield x and y coordinates on the target projected on the screen. The main performance measure was the shot group precision. The "center of mass" of the first N shots was calculated initially and the shot group precision was defined as the mean dispersion across N shots to the center of the shot group. Two dimensional standard deviation (x & y axis) of each shot from the center was also recorded.

Incorporating all sensor inputs, the PPT setup provides direct, real-time haptic feedback to the shooter regarding their pre-shot alpha level (in relation to their baseline level) and HR, as well as real-time continuous measures of respiration, trigger pressure, and muzzle wobble in relation to shots. In this manner, the PPT setup is designed to provide feedback not only regarding the output of their comprehensive technique (shot performance), but also feedback on each element of technique that contributes to shot performance.

The PPT lab setup is currently being developed as a generic training/research platform that could be extended to many other activities requiring skill development. It is an ideal solution for indoor training; however the multiple display devices such as the computer and the LaserSoft rifle simulation make it bulky and not easily portable. Based on the results from the lab setup, the most relevant subset of features was implemented in a field-deployable version which we call the "PPT-mobile". We developed one such version that extracted EEG alpha level as well as EKG R-wave spikes in real-time and provided feedback on alpha level and heart rate via haptics and audio. The same ARM9 microprocessor that was used in the lab setup was also used in the mobile version, and the EEG / EKG sensors were interfaced to it. A Bluetooth module was also interfaced to collect data on a remote computer if required. The device was battery operated and was attached to the EEG sensor cap worn comfortably by the user. The device had a modular architecture allowing for easy addition of new features.

2.2 Pilot Study

We recruited 9 novice subjects (8 males, 1female; mean age 23.1 years; range 20-27) to evaluate the effectiveness of the real-time PPT when applied to marksmanship training. Three participants first completed the protocol without sensory feedback (no PPT) and then returned on a later day to complete the PPT protocol. The remaining six subjects completed the PPT protocol only. Preliminary studies suggested that sensor-based feedback is overwhelming to the novice shooter in the beginning training stages, and is more useful once the novice is familiar with the positional elements of marksmanship. For this reason, real-time feedback using the PPT during shooting was not delivered to the novice until the later trials of the study, after they had received computer instruction, individualized coaching, and modular training in EEG alpha, breath, and trigger control.

As a preliminary assessment of the efficacy of the PPT, performance improvements achieved using the PPT were compared to two other experimental groups of novices, and one group of marksmanship experts. All novice groups (Ground Zero, Tx2, and PPT) represented similar age and experience levels, completed the same number of trials and watched the same 15-minute introductory marksmanship video. All other factors being equal, each novice group completed marksmanship training with distinct conditions. The Ground Zero group received no individualized coaching, no offline (not while shooting) sensor-based feedback, and no PPT while shooting. The Tx2 group received individualized coaching and offline sensor-based feedback, but no PPT while shooting. The PPT group received individualized coaching, offline sensor-based feedback, and PPT while shooting. The expert group was comprised of 10 USMC marksmanship coaches recruited from USMC base Camp Pendleton, each of which qualified as expert on their most recent marksmanship qualification. Experts were given no instruction or other type of feedback during shooting.

3 Results

Preliminary analysis of performance data comparing the performance (mean distance of each shot from shot group center) at Baseline (Trials 1 and 2) and Final (two trials

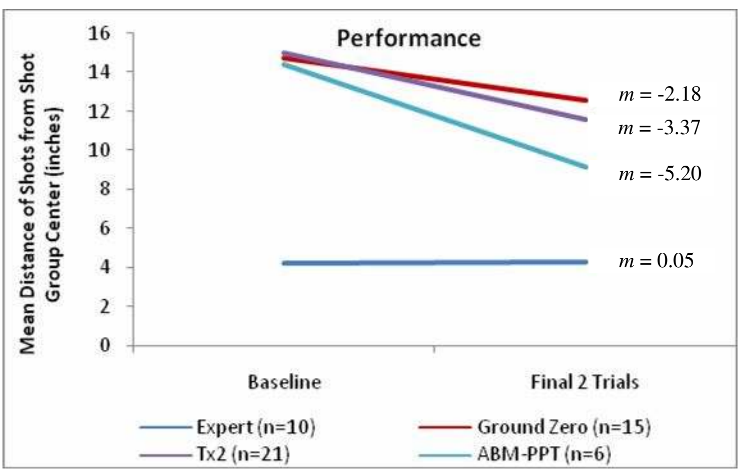


Fig. 2. Performance curve of groups trained with & without PPT

at or after Trial 7 that show best performance) of PPT protocol compared to the same measures for other experimental groups revealed an improved learning trajectory beyond that attained with the Ground Zero or Tx2 groups. The performance data (Fig. 2) is encouraging and suggests that providing offline, real-time physiological feedback through the PPT protocol is effective in improving the performance of novices at a greater rate than for other treatment groups.

4 Discussion

The PPT is a part of our effort to create a suite of Interactive Neuro-Educational Technologies (I-NET) that can be used in multiple training environments. I-NET covers four major themes: 1) integrating brain monitoring into paced instructional tutorials, 2) identifying psychophysiological characteristics of expertise using expert marksman as a model population, 3) developing sensor-based feedback to assist novices in acquiring marksmanship skills and 4) identifying neurocognitive factors that are predictive of marksmanship skill acquisition to allow early triage and interventions. While the entire PPT system is limited in its functionality due to the reliance on computer monitors and projectors to display visual feedback, the development of the PPT-mobile opens the door for incorporation of real-time neuro- and biofeedback for the first time into many psychomotor skill tasks performed outside of the laboratory setting.

Experiments are currently underway to begin to address the applicability of the PPT-mobile in scenarios that more closely approximate combat conditions. It will be necessary to introduce threat/fear stressors into the basic marksmanship training setup to encourage a combat mindset, as being able to manage such stress is critical. One way in which this will be implemented is through the use of first-person shooter games. The psychophysiological profile of expert and novice shooters under stressful combat conditions is as yet undefined; mitigation is usually performance driven, and via intense reality-based training. Thus, developing neurophysiological metrics that drive sensor-based feedback is highly desirable. In the current (non-combat) training paradigm, the measure of shooter readiness that drives the PPT is based on high alpha + HR deceleration. How this measure relates to performance under more dynamic conditions is unclear, particularly when there is a threat component.

Multiple combinations of pre-training triage and interventions including relaxation and attentional training will be evaluated in addition to real-time feedback during simulated combat marksmanship. The psychophysiological profile of the expert USMC marksmen suggests a finely tuned level of control over physiology that appropriately allocates resources to meet task demands [31, 32]. I-NET are being designed to assess, characterize and further develop this psychophysiological control system. The acquisition of expertise in marksmanship can serve as a model of the key skills required for training in military and other educational environments and can be extended to other activities such as golf, archery and free throw shooting in basketball.

Acknowledgements

This work was supported by the DARPA program, Accelerated Learning, Dr. Amy Kruse Program Manager and by NIH grant MH078436.

References

1. Fisk, A.D., Schneider, W.: Memory as a function of attention, level of processing, and automatization. J. Exp. Psychol. Learn Mem. Cogn. 10(2), 181–197 (1984)
2. Schneider, W., Shiffrin, R.M.: Controlled and automatic human information processing I: Detection, search, and attention. Psychological Reviews 84, 1–66 (1977)
3. Schneider, W., Fisk, A.D.: Concurrent automatic and controlled visual search: Can processing occur without cost? Journal of Experimental Psychology: Learning, Memory, and Cognition 8, 261–278 (1982)
4. Anderson, J.R.: Acquisition of cognitive skill. Psychological Review 89, 369–406 (1982)
5. Anderson, J.R., Cognitive psychology and its implications. 4 ed. 1995, New York: Freeman.
6. Segalowitz, S., Wintink, A., Cudmore, L.: P3 topographical change with task familiarization and task complexity. Cognitive Brain Research 12(3), 451–457 (2001)
7. Klimesch, W.: EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res. Rev. 29(2-3), 169–195 (1999)
8. Chang, G., et al.: Practice-related EEG Changes during Simulated Flight. In: MIT Conference, MIT Press, Cambridge (2003)
9. Crews, D., Lander, D.: Electroencephalographic measures of attentional patterns prior to the golf putt. Medicine and Science in Sports and Exercise 25, 116–126 (1993)
10. Landers, D., Han, M., Salazar, W.: Effects of learning on electroencephalographic and electrocardiographic patterns in novice archers. International Journal of Sports Psychology 25, 313–330 (1994)
11. Kerick, S., Douglass, L., Hatfield, B.: Cerebral cortical adaptations associated with visuomotor practice. Medicine and Science in Sports and Exercise 36(1), 118–129 (2004)
12. Haufler, A., et al.: Neuro-cognitive activity during a self-paced visuospatial task: comparative EEG profiles in marksmen and novice shooters. Biological Psychology 53(2-3), 131–160 (2000)
13. Hillman, C., et al.: An electrocortical comparison of executed and rejected shots in skilled marksmen. Biological Psychology 52(1), 71–83 (2000)
14. Hatfield, B., et al.: Electroencephalographic studies of skilled psychomotor performance. Journal of Clinical Neurophysiology 21, 144–156 (2004)
15. Kerick, S., Hatfield, B., Allender, L.: Event-related Cortical Dynamics of Soldiers during Shooting as a Function of Varied Task Demand. Aviation Space and Environmental Medicine (in press)
16. Kerick, S., et al.: The role of the left temporal region under the cognitive motor demands of shooting in skilled marksmen. Biological Psychology 58(3), 263–277 (2001)
17. Loze, G., Collins, D., Holmes, P.: Pre-shot EEG alpha-power reactivity during expert air-pistol shooting: a comparison of best and worst shots. Journal of Sports Sciences 19(9), 727–733 (2001)
18. Tremayne, P., Barry, R.J.: Elite pistol shooters: physiological patterning of best vs. worst shots. International Journal of Psychophysiology 41(1), 19–29 (2001)
19. Salazar, W., et al.: Hemispheric asymmetry, cardiac response, and performance in elite archers. Research Quarterly for Exercise and Sport 61(4), 351–359 (1990)
20. Kontinnen, N., Lyytinen, H., Viitasalo, J.: Preparatory heart rate patterns in competitive rifle shooting. Journal of Sports Sciences 16(3), 235–242 (1998)
21. Proctor, R., Dutta, A.: Skill Acquisition and Human Performance. Sage Publications, Thousand Oaks (1995)

22. Feltz, D., Landers, D.: The effects of mental practice on motor skill learning and performance: a meta-analysis. Journal of Sport Psychology 5, 25–57 (1983)
23. Hatfield, B., Hillman, C.: The psychophysiology of sport: A mechanistic understanding of the psychology of superior performance. In: Singer, R., Hausenblas, H., Janelle, C. (eds.) Handbook of Sports Psychology, pp. 362–386. Wiley & Sons, New York (2001)
24. Berka, C., et al.: Objective Measures of Situational Awareness using Neurophysiology Technology. In: Schmorrow, D., Stanney, K., Reeves, L. (eds.) Augmented Cognition: Past, Present and Future, pp. 145–154. Strategic Analysis, Inc., Arlington (2006)
25. Stevens, R., Galloway, T., Berka, C.: Integrating EEG Models of Cognitive Load with Machine Learning Models of Scientific Problem Solving. In: Schmorrow, D., Stanney, K., Reeves, L. (eds.) Augmented Cognition: Past, Present and Future, pp. 55–65. Strategic Analysis, Inc., Arlington (2006)
26. Berka, C., et al.: Evaluation of an EEG-Workload Model in an Aegis Simulation Environment. In: Caldwell, J.A., Wesensten, N.J. (eds.) Proceedings of SPIE Defense and Security Symposium, Biomonitoring for Physiological and Cognitive Performance during Military Operations, pp. 90–99. SPIE: The International Society for Optical Engineering, Orlando (2005)
27. Berka, C., et al.: Real-time Analysis of EEG Indices of Alertness, Cognition, and Memory Acquired with a Wireless EEG Headset. International Journal of Human-Computer Interaction 17(2), 151–170 (2004)
28. Best, J.B.: Cognitive Psychology, 3rd edn. West Publishing Company, New York (1992)
29. Solomon, R.M., Horn, J.M.: Post-shooting traumatic reactions: A pilot study. In: Reese, J., Goldstein, H. (eds.) Psychological Services, in Law Enforcement, pp. 383–393. United States Government Printing Office (1986)
30. LaserShot, Military Skills Engagement Trainer. Stafford, Tx (2008)
31. Berka, C., et al.: Using Interactive Neuro-Educational Technology to Increase the Pace and Efficiency of Rifle Marksmanship Training. American Educational Research Association, New York (2008)
32. Pojman, N., Behneman, A., Johnson, R., Berka, C., Chung, G., Nagashima, S., Espinosa, P.: Identifying Psychophysiological Metrics of Expert Marksman. In: 13th International Conference on Human-Computer Interaction, San Diego (2009)

The Quality of Training Effectiveness Assessment (QTEA) Tool Applied to the Naval Aviation Training Context

Tom Schnell[1], Rich Cornwall[2], Melissa Walwanis[3], and Jeff Grubb[4]

[1] Operator Performance Laboratory (OPL), 3131 Seamans Center, University of Iowa, Iowa City, Iowa 52242
[2] Alion Science and Technology, BMH-Operation, 5365 Robin Hood Road Suite 100, Norfolk, VA 23513
[3] Naval Air Warfare Center, Training Systems Division, AIR -4651, Training & Human Performance R&D Branch, 12350 Research Pkwy Orlando, FL 32826-3275
[4] Naval Aerospace Medical Institute, Pensacola, FL 32501
thomas-schnell@uiowa.edu, rcornwall@alionscience.com, melissa.walwanis@navy.mil, Jefferson.Grubb@med.navy.mil

Abstract. Today, flight trainers use objective measures of task performance and additional estimated, subjective data to assess the cognitive workload and situation awareness of trainees. This data is very useful in training assessment but trainees can succeed at performing a task purely by accident (referred to as "miserable success"). Additionally the trainee can be in a less than optimal for learning cognitive state when the instructor operator applies brute force training tasks and methods with little regard to the learning curve which can result in the training being too easy or, more often, too difficult, thereby inducing negative learning. In order to provide the instructor with additional quantitative data on student performance, we have designed the Quality of Training Effectiveness Assessment (QTEA) concept. QTEA is conceived as a system that allows the trainer to assess a student in real-time using sensors that can quantify the cognitive and physiological workload.

Keywords: Neurocognitive measures, operator state characterization, flight training.

1 Introduction

QTEA is based on a technology called the Cognitive Avionics Tool Set (CATS) that has been developed over the last 4-5 years at the Operator Performance Laboratory (OPL). CATS uses neurocognitive and physiological signals of the pilot to generate a real-time measure of cognitive workload. CATS has been tested in flight simulators, research vehicles, and flight test aircraft (Bell 412 at the National Research Council in Canada and BE-36 at OPL). QTEA is conceived as a system that allows the trainer to assess a trainee in real-time using sensors that can quantify the cognitive and physiological workload. Using QTEA, the trainer can quantify the student's workload level

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 640–649, 2009.

in real-time so that the scenarios can be adjusted to an optimal intensity. The cognitive and physiological measures also serve as a quantitative manifestation of a student's learning curve and it will be possible for the trainer to detect plateaus in learning. Using QTEA, the trainer will be able to assess the needs for further training in a student. Using this real-time data, training scenarios can be optimally adapted in terms of difficulty to maximize the effectiveness of learning. This will likely reduce training costs by reducing training times and may increase operational success levels.

By using physiologically based measures to corroborate objective and subjective measures, trainers will have a quantitative tool to measure training effectiveness. Our approach is based on using a battery of neurophysiological and physiological sensors on the trainee and to fuse this data with mission data, apply sophisticated signal processing and classification techniques, to gain a fuller picture of training effectiveness. The basic idea of QTEA is to give the trainer a real-time picture of the performance of a trainee based on human physiological and cognitive data, flight technical, and mission specific data. QTEA is a sensor-fused, real-time performance gauge that will allow training systems to optimize training scenarios in real-time and provide superior information for after-action review.

Today, trainers do not have access to quantitative cognitive and physiological data of their trainees in real-time and they have to subjectively judge cognitive loading through observation. For optimal training effectiveness, it is imperative that trainees are exposed to scenarios with optimal difficulty levels. In some cases, this means pushing the trainee to the edge of cognitive capacity that can be quantified using QTEA. This tool will provide the trainer with real-time knowledge of trainee attention and cognitive bandwidth so that scenarios can be tuned in real-time to achieve maximum transfer of training. Since QTEA characterizes behavioral data, it also portrays the learning curve of a trainee. For example, the eye tracking sensor shows the trainer how the trainee is acquiring visual information, allowing for immediate intervention and remediation if poor scan techniques are detected. This prevents formation of bad habits in visual scanning. QTEA can tell the trainer if the trainee is cognitively overtaxed by certain tasks, thus providing instant feedback about workload optimization.

2 Components of QTEA

The QTEA system consist of sensors that can be installed in the training environment and/or deployed on the trainee to monitor neurophysiological and physiological measures. A processor (PC computer) is needed to perform the computational functions such as sensor interfacing, signal processing, operator state classification, and interfacing with client applications as indicated in Fig. 1. The processor is a normal PC based computer and is a part of the final QTEA product. In addition, QTEA consists of software that interfaces with the sensors, synchronizes the data with mission specific performance data, characterizes operator state, and exports that data throughout the High Level Architecture (HLA, IEEE 1516) for consumption in scenario management tools such as the Common Distributed Mission Training Station (CDMTS) that was selected as the integration platform because it is the controller station where the trainer has direct influence on the scenarios and knowledge of results (KR).

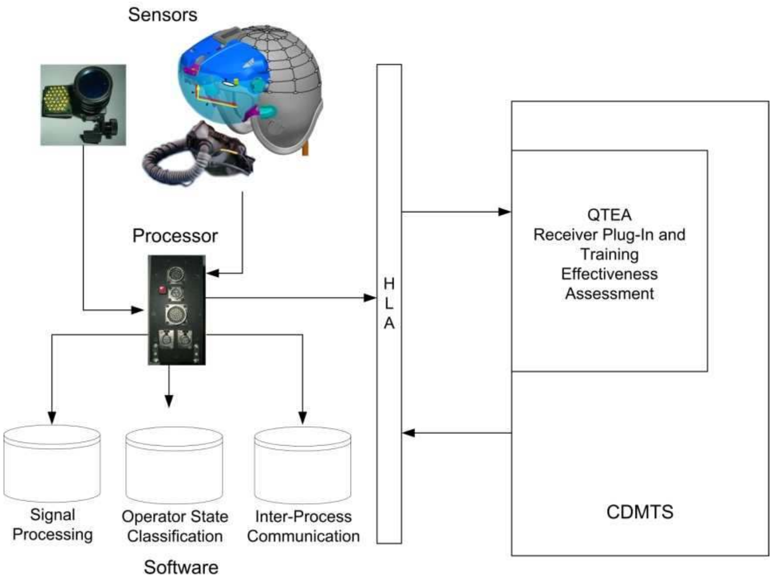


Fig. 1. Components of QTEA

Eitelman [1] pointed out the suitability of CDMTS to be enhanced with AugCog like technology. QTEA will feature a component that serves as a plug-in into client applications such as CDMTS where it will integrate the operator state data with other performance data to express the quality of training effectiveness. QTEA is not dependent on CDMTS and will be able to connect to any other HLA capable application that requires operator state data. The neurocognitive and physiological sensor system in QTEA is called the Cognitive Avionics Tool Set (CATS). CATS has been discussed by Schnell et al. [2][3][4][5] and was tested in several airborne and automobile applications. CATS evolved from the architecture used in the Computerized Airborne Research Platform (CARP) flight test aircraft at the OPL. The central element of CATS is the Input-Output-System (IOS) computer. Data from all sensors onboard the aircraft generate streams that are captured as asynchronous busses in the IOS. Our sensors typically have intelligent pre-processors that perform signal conditioning functions and the analog-digital conversion. These pre-processors communicate with the IOS by means of Ethernet packets. It should be noted that in our paradigm, we use human sensor data and (aircraft) system sensor data to classify operator state. This means that not only data from the physiological sensors is used to determine the state of the operator, but also data from the air data computer (ADC), Attitude Heading Reference System (AHRS), Global Positioning System (GPS), Inertial Navigation System (INS), system switches, and mission data. Thus, we use physiological data, flight technical data, and mission specific data to determine operator state in a robust fashion.

2.1 QTEA Sensors

The sensors used in QTEA can measure a wide range of human neural and physiological characteristics. It is important to note that sensors alone do not facilitate operator classification. The sensor signals must be properly filtered to eliminate environmentally and physiologically induced noise and artifacts. Processing algorithms must be employed to turn raw data into meaningful metrics. The effort involved in properly filtering and processing signals is complex and time-consuming, but is extraordinarily important in order to effectively classify. We have conducted extensive literature reviews [2][3][4][5] to assess strategies employed by other researchers. The CATS framework easily accommodates new sensors and analysis techniques into the classification network. Our guiding design philosophy is to initially deploy QTEA with a battery of sensors that have a technical readiness level (TRL[1]) of at least 6 (see Table 1). We then intend to refine lower TRL sensors to the same or higher level. The sensors that we are focusing on include eye tracking (estimated TRL 7), electrocardiogram (ECG, estimated TRL 7), respiration (estimated TRL 6), galvanic skin response (GSR, estimated TRL 6), and electroencephalogram (EEG, estimated TRL 5).

Pilots need to visually scan different sources of information (instruments, outside, etc.) and integrate this information into situation awareness and situational evidence that will support motor control and decision making. The eye scan measurement system that we are using can accommodate one to six remote mounted cameras to track facial features, head position (6 dof), and to generate a binocular gaze vector in a true 3D environment. CATS generates eye scanning measures that include fixation duration, fixation frequency, saccade velocity, saccade length, dispersion measures, and link analyses between areas of interest (AIOs). These measures can then be correlated with workload by means of a regression model.

When neurons fire in the brain, they generate electrical pulses that radiate outward to the scalp. Electroencephalogram (EEG) measures the electrical potential along the scalp to quantify the level of activity in the brain. We use a high density, 128-channel sensor at a high update rate (1000 Hz). Our net provides high spatial and temporal resolution. The CATS software suite has an extensive set of tools to manage channel integrity, detect and suppress artifacts and provide PSD and clinical band analysis. The EEG power levels in the different spectral bands are used in CATS to generate measures of effectiveness that are regressed against workload using a regression model.

We also use a three channel electrocardiogram (ECG) sensor to measure cardiac electrical activity. The ECG strongly reflects sympathetic and parasympathetic activity in the autonomous nervous system. Heart rate alone is affected by several factors,

---

[1] TRL 1 Basic principles observed and reported, TRL 2 Technology concept and/or application formulated, TRL 3 Analytical and experimental critical function and/or characteristic proof-of concept, TRL 4 Component and/or breadboard validation in laboratory environment, TRL 5 Component and/or breadboard validation in relevant environment, TRL 6 System/ subsystem model or prototype demonstration in a relevant, TRL 7 System prototype demonstration in a flight environment, TRL 8 Actual system completed and "flight qualified" through test and demonstration, TRL 9 Actual system "flight proven" through successful mission operations.

making it less valuable for assessing autonomic activity. However, heart rate variability (HRV), or the change in heart rate from beat to beat, provides strong indication of autonomic activity. Several analytical methods have been applied to the analysis of HRV including the discrete Fourier transform, the discrete wavelet transform, Poincare analysis. The Puls Atem Quotient which compares the heart rate to breathing rate also provides an indication of stress. The heart exhibits a unique electrical pattern known as the PQRST waveform during each beat. Changes in the PQRST waveform from beat to beat can also be used to assess operator state.

A pulse-oximeter measures oxygen levels in the bloodstream. We measure pulse-oxigenation on the forehead. Measuring blood oxygen levels is useful for detecting the onset of hypoxia during flight. The pulse-ox sensor can also be used to derive heart rate. Respiration is measured through a flexible belt worn around the chest, or by virtue of gas-flow sensors integrated in an aviator oxygen system. Respiration rate and tidal volume can be measured. Respiration provides some information about autonomic nervous activity. Respiration modulates heart rate variability (HRV). Effective analysis of HRV requires knowledge of respiration to correct for the modulation transformation.

We are currently adapting this capability of QTEA for use in a reduced Oxygen Breathing Device (ROBD) application. The Reduced Oxygen Breathing Device is a training system that provides mixed gas through an aviator's oxygen mask for hypoxia recognition and recovery training. The ROBD simulates hypoxia by diluting room air with nitrogen, resulting in lower oxygen content. The ROBD can simulate altitudes in excess of 30,000 ft, with altitudes and profiles selected by the instructor. The ROBD can be paired with a simple flight simulator so that pilots can experience the symptoms of hypoxia while performing flying tasks. Additionally, ROBD training will be provided to Naval Flight Officers. In this case, a simulation of NFO tasks will be provided and the trainer networked so that the NFO and pilot can fly the same mission and practice proper crew communication during the hypoxia scenario. Using physiological performance measures, the trainers can thus understand how trainee flight performance deteriorates as a function the simulated altitude level under increasing hypoxia.

2.2 QTEA Performance Metrics

There are two types of performance measures in QTEA, mission specific and physiological ones. The mission specific measures will change with the use case of QTEA. We have selected the close air support (CAS) task using non-precision weapons. This task is very demanding in terms of timing, planning, and aircraft control during delivery of the weapon. For naval aviation in a Close-Air-Support (CAS) task, the measures of effectiveness include airmanship measures (flight technical), administrative, and tactical. The physiological measures describe the body's reaction to the task and include cognitive measures quantifying stress, cognitive resources, and attention (alertness). The physiological measures subsystem of QTEA functions automatically and in real-time, providing the trainer with a metric that indicates cognitive loading. The performance metrics are organized in vectors describing the human (pilot), the aircraft, and the environment. A regression model at the heart of the classifier accepts multiple input vectors from which it classifies operator workload (Fig. 2.). No single

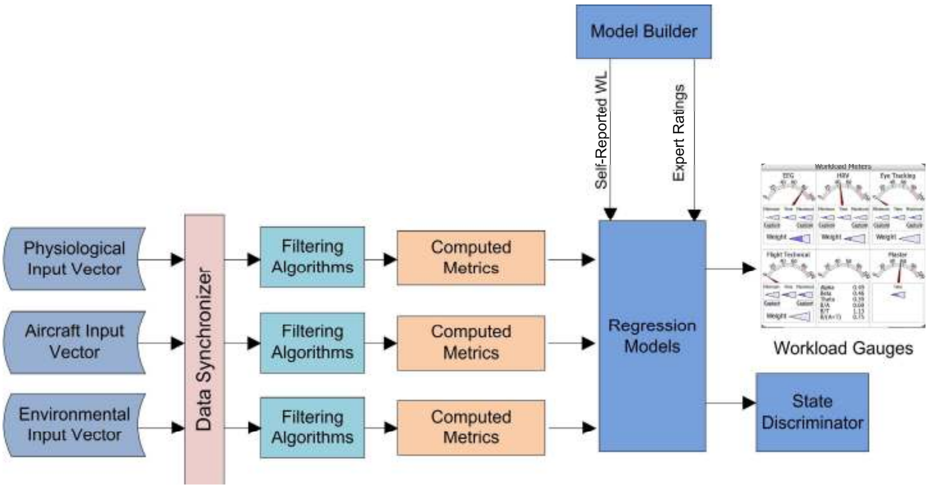


Fig. 2. Simplified Flow of Input Vectors from Sensors to Classifier

vector alone provides the complete picture. A true understanding of the situation requires an understanding of all of the input vectors. It is only from this composite picture that the classifier is able to assess operator state. We implemented the multi-sensory state classification system in CATS using a regression model that was developed on the basis of empirical data.

3 Demonstration of Workload Measurement with CATS/QTEA

To test the ability of the QTEA prototype to measure workload and to demonstrate its usability in a distributed training simulation environment, we used the existing Live Virtual Constructive (LVC) federation shown in Fig. 3 that is available at OPL. This framework consists of a live component in the form of a highly modified Aerovodochody L-29 fighter jet trainer called the Cognitive Delfin (COD), two networked flight simulators (full flight deck simulator and fast jet simulator), and constructive entities generated by the Joint Semi-Automated Forces (JSAF) system, controlled through CDMTS. Sector and forward controllers can participate with respective applications in this federation through communication over a TeamSpeak channel. The live COD is connected to the federation through a long range data link using a 900 MHz spread spectrum radio with a rotator antenna on the ground station. The ground station is connected to the simulator network through an Ethernet connection. The pilot in the rear seat crew station is the exercise participant (usual call sign is Herkey 11). The test director (call sign OZ) is using a Rockwell Collins EPX-500 visualization system to maintain overall situation Awareness (SA) of all entities in play and two TeamSpeak channels (exercise communications and stealth communication) to coordinate the exercise with all participants.

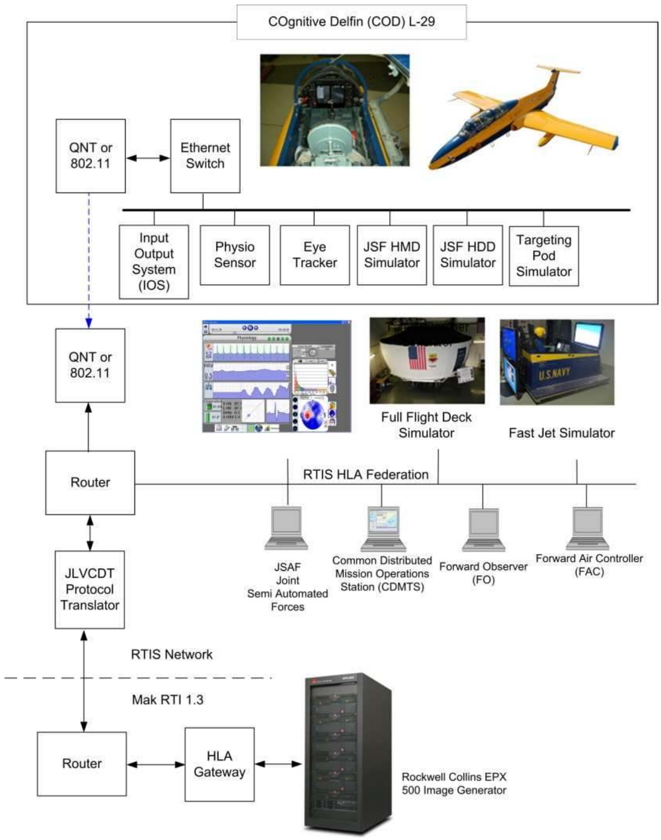


Fig. 3. Live Virtual Constructive Framework for Testing Workload

For this demonstration, neurocognitive (EEG), physiological (ECG, GSR, Respiration), and eye tracking sensors were deployed on the pilot flying the fast jet simulator only. Workload is classified by CATS in several sub-categories including cognitive (EEG_, physiological (primarily HRV), eye tracking (saccade and fixation durations), and flight technical (flight technical errors). The workload scores are exported by CATS to HLA.

The CAS scenario involved two F/A-18s on patrol near Solvang, California, being requested to eliminate tanks on the Santa Inez River Bridge (red triangle at river in Fig. 4.).

Although the two jets were virtually on patrol over the California countryside, in reality, one of the jets, using the call sign Herkey 11, was a pilot in the rear seat of the COD, flying in real airspace near Iowa City, Iowa. The second asset, using the call sign Herkey 12, was a pilot in the fast jet simulator. Both the live (COD) and virtual (fast jet simulator) pilots could see the virtualized terrain, and standard radio communication was used to call the mission. The mission started with a call for fire from a forward air controller (FAC) with the call sign OBAN. The FAC relayed the call for

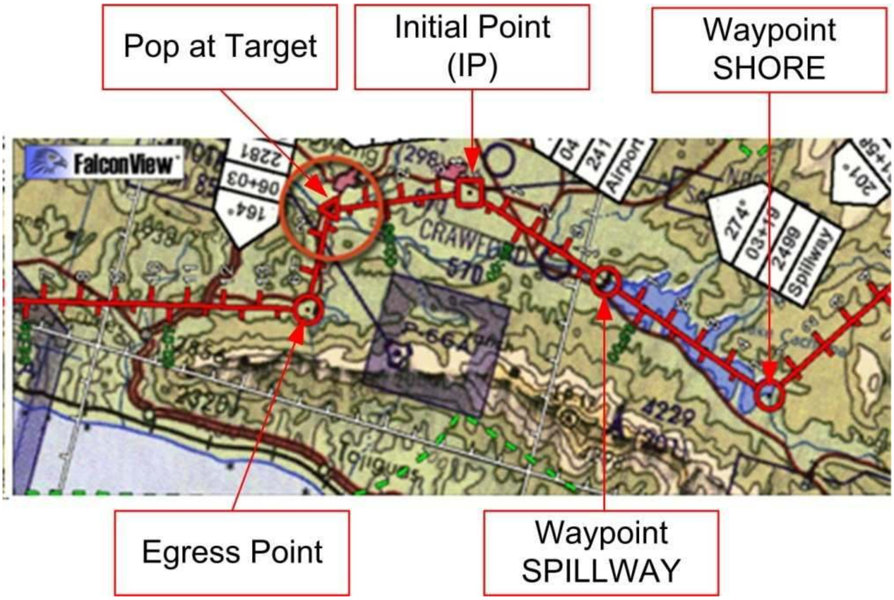


Fig. 4. LVC Close Air Support Mission: Defense of Santa Inez River Bridge

fire to a sector controller with the call sign TALISKER. Both controllers were physically located at the central command and control station at Rockwell Collins in Cedar Rapids, Iowa. TALISKER located the two F/A-18s (Herkey flight) in a holding area NE of the target. A route was generated and communicated to Herkey flight. The route started at the flight's current location in the holding pattern at CAP. Subsequent waypoints were called SHORE, SPILLWAY, AIRPORT (IP), TARGET, and GTFO. TALISKER cleared Herkey 1 onto the route and handed communications off to OBAN who tasked Herkey flight with a 9-line briefing. Herkey 11 (back seater in COD) copied the 9 line and Herkey 12 (fast jet pilot) acknowledged the target assignment. Herkey 11 followed the flight plan and Herkey 12 followed in combat spread using the outside visualization in the simulator. The virtualized target (bridge in Solvang) was set up so that it coincided with a bridge in a similar orientation in Iowa. That way, it was possible for Herkey 11 to acquire the target visually. Both Herkeys used a popup delivery profile to attack their assigned targets and after the attack they cleared six and formed up for egress to GTFO at which point the simulation terminated (SIMEX). Data was recorded with CATS and real-time workload scores were broadcast onto HLA for use in CDMTS.

In order to illustrate the results of the demonstration, we tagged the workload data by waypoint as shown in Fig. 5. The EEG scores appear to react very quickly to the increased task demand of the mission, whereas the HRV scores appear to lag from 30 seconds to a minute. Following the waypoints from CAP to AIRPORT is easy from a flight technical point of view and this is well reflected in the low EEG workload scores. As the flight approaches the IP the workload goes up as the pilot starts to focus on targeting and visually identifying the target. The actual popup attack is the

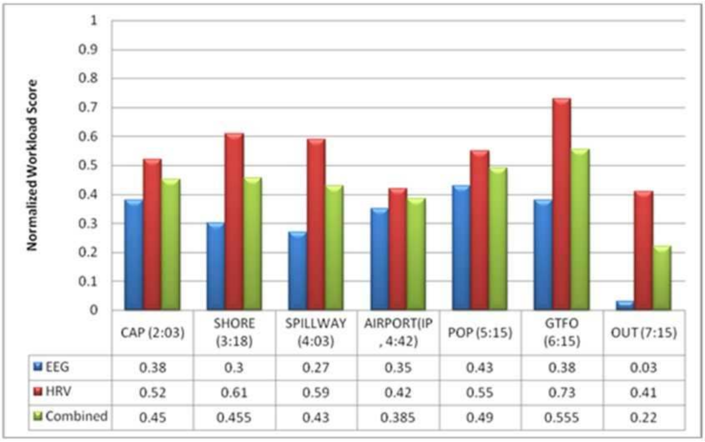


Fig. 5. Workload Scores for Close Air Support Task

most demanding portion of the flight, and the EEG scores appear to reflect that demand. The EEG workload score decreases after the pop at TARGET and drops significantly during cruise egress to waypoint OUT. The HRV score was relatively high during the waypoint sequence and then actually dropped when reaching the IP, increased during the pop at TARGET and was highest right after the pop during egress to the nearby waypoint GTFO. It seems that the HRV data has a considerable delay on the order of 20 seconds. Thus, the bars for the HRV at each waypoint more accurately describe the workload level that about a minute prior.

4 Conclusions

Instrumented aviation training systems such as QTEA provide instructors with quantitative data of the student's performance. This data can be used by automated scenario generation systems to adjust scenario intensity in real-time to maximize learning by keeping stimulation at its optimal level. The quantitative data generated by QTEA also provides for superior after-action review, offering the instructor and student the ability to review deviations in mission or flight technical domains as well as the occurrence of cognitive (workload) bottlenecks, poor control manipulation, or ineffective eye scanning technique. Through review and discussion of such quantitative data, the instructor and student can develop training strategies that achieve the training goal in a shorter time than would be possible without such advanced tools. With its neurocognitive and physiological battery of measures, in conjunction with flight technical and mission specific performance measures, QTEA is able to characterize the training patterns of experts and compare them to patterns of novices. By comparing generic portions of those "expert" patterns with the patterns obtained by a trainee as he/she progresses through the learning curve, we expect to have a quantitative assessment of the quality of training. The simulator training community could benefit from quantitative tools such as QTEA that measure the effectiveness of training on the basis of human performance data. Using the real-time workload gauge in QTEA, it should be possible to adapt training scenario difficulty to maximize the effectiveness of learning. This will likely reduce training cost by reducing training time.

References

[1] Eitelman, S.M., Wheeler-Atkinson, B.F., Walwanis-Nelson, M., Stitso, M.E.: Enhancing the Common Distributed Mission Training Station Through Augmented Cognition Technology. In: Proceedings of the Interservice/Industry Training, Simulation & Education Conference, I/ITSEC (2006) (Conference Theme: Training the 21st Century Joint Force)

[2] Schnell, T., Keller, M., Macuda, T.: Application of the Cognitive Avionics Tool Set (CATS) in airborne operator state classification. In: Augmented Cognition International Conference, Baltimore, MD (October 2007a)

[3] Schnell, T., Keller, M., Macuda, T.: Pilot state classification and mitigation in a fixed and rotary wing platform. In: Aerospace Medical Association (ASMA) annual conference, New Orleans, LA (April 2007b)

[4] Schnell, T., Macuda, T., Keller, M.: Operator state classification with the cognitive avionics tool set. In: Schmorrow, D., Stanney, K. (eds.) Augmented Cognition: A practitioner's guide (2008)

[5] Schnell, T., Macuda, T., Poolman, P.: Toward the cognitive cockpit: Flight test platforms and methods for monitoring pilot mental state. In: Augmented Cognition International Conference, San Francisco, CA (October 2006)

[6] Roscoe, A.H.: Assessing Pilot Workload in Flight. Flight Test Techniques. In: Proceedings of NATO Advisory Group for Aerospace Research and Development (AGARD), Neuilly Sur Seine, France (1984)

Perceptually-Informed Virtual Environment (PerceiVE) Design Tool

Anna Skinner[1], Jack Vice[1], Corinna Lathan[1], Cali Fidopiastis[2], Chris Berka[3], and Marc Sebrechts[4]

[1] AnthroTronix, Inc. 8737 Colesville Rd., L203 Silver Spring, MD 20910
[2] Institute for Simulation and Training, University of Central Florida 3100 Technology Parkway, Orlando, FL 32826
[3] Advanced Brain Monitoring, Inc. 2237 Faraday Ave., Ste 100 Carlsbad, CA 92008
[4] Catholic University of America, Department of Psychology, 620 Michigan Ave, NE Washington, DC 20064
{askinner,jvice,clathan}@atinc.com, cfidiopia@ist.ucf.edu, chris@b-alert.com, sebrechts@cua.edu

Abstract. Virtual environments (VE's) are becoming more and more prevalent as training tools for both military and civilian applications. The common assumption is that the more realistic the VE, the better the transfer of training to real world tasks. However, some aspects of task content and fidelity may result in stronger transfer of training than even the most high fidelity simulations. This research effort seeks to demonstrate the technical feasibility of a Perceptually-informed Virtual Environment (PerceiVE) Design Tool, capable of dynamically detecting changes in operator behavior and physiology throughout a VE experience and comparing those changes to operator behavior and physiology in real-world tasks. This approach could potentially determine which aspects of VE fidelity will have the highest impact on transfer of training. A preliminary study was conducted in which psychophysiological and performance data were compared for a visual search tasks with low and high fidelity conditions. While no significant performance effects were found across conditions, event-related potential (ERP) data revealed significant differences between the low and high fidelity stimulus conditions. These results suggest that psychophysiological measures may provide a more sensitive and objective measure for determining VE fidelity requirements.

Keywords: Psychophysiological Measures, Virtual Environments, Fidelity, Transfer of Training, Simulation Design.

1 Introduction

Virtual environments (VE's) and simulations are being employed for training applications in a wide variety of disciplines, both military and civilian. Technological advances are enhancing the ability of developers to create VE's with visual, auditory, haptic, and even olfactory realism. Such VE's allow the military to train skills that are too costly, too dangerous, or are otherwise impossible to practice. While a significant

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 650–657, 2009.

research has been conducted examining the transfer of training from VEs (for example, [1, 2]), only a limited number of efforts have used psychophysiological measures to do so. The common assumption is that the more realistic the VE, the better the transfer of training to real world tasks. However, some fidelity components (e.g., display resolution, frame rate, texture mapping, physics modeling, etc.) may result in stronger transfer of training than others for a given task or domain. This has traditionally been determined by performance measurements compared before and after design iterations. With each design modification, end users are tested using the VE and their performance is compared to performance on the prior VE design. Improved performance is often assumed to be related to improved design and fidelity. However, it is difficult to identify the specific design components that directly relate to transfer of training improvements. Furthermore, this method of design focuses on trial and error, and is therefore time consuming, undirected, and may result in false associations between performance and VE characteristics. For example, unless each component of the new simulator design is introduced separately, it will not be known which fidelity design improvements bear the strongest significance to performance improvements. Thus, a more sensitive, objective, and comprehensive assessment of the quality of interaction with a simulation is needed to effectively identify the specific components of simulation that bare relevance to real world operational tasks.

One of the major questions simulation designers must address is "what components of fidelity have the greatest impact on transfer of training?" Fidelity is defined as the degree to which features (e.g., visual, auditory, etc) in the Virtual Environment (VE) match features in the real environment. Following this premise, one can argue that a VE with maximum fidelity would result in transfer of training equivalent to real-world training since the two environments would be impossible to differentiate [3; Martin, 1981). However, developers are limited by practical restrictions such as cost, time, and development resources. Thus, trade-offs are necessary. There is currently a limited understanding of the specific trade-offs between increases in simulation fidelity and operator behavior, and essentially no guarantee to developers that a particular level/area of simulation fidelity is sufficient to provide effective transfer of training.

Under an Office of Naval Research-funded Small Business Technology Transfer (STTR) effort the authors proposed to develop a Perceptually-informed Virtual Environment (PerceiVE) Design Tool, which utilizes physiological measures to determine fidelity requirements with the goal of optimizing transfer of training between simulated and real world tasks. We hypothesized that a physiologically-based system capable of dynamically detecting changes in operator behavior and physiology throughout a VE experience, and comparing those changes to operator behavior and physiology in real-world tasks, could potentially determine which aspects of VE fidelity will have the highest impact on transfer of training.

EEG and event related potential (ERP) approaches offer excellent temporal resolution for tracking of neural activity representing the flow of information from sensory processing, detection and identification of relevant objects, and decision-making. ERP signature components associated with the identification of target stimuli were first reported in 1965 and named "P300s or P3b or Late Positivity" [4, 5], (Squires, Squires, & Hillyard, 1975) because target stimulus presentations are associated with large positive potentials maximal over parietal cortex with peak latency ranging from 300-800 ms after presentation of the target stimulus. The P300 is generally accepted to be a post-sensory signal elicited when subjects attend and respond to target stimuli and is believed to be related to higher cognitive processes including updating working memory [5]. Several reports suggest that when target stimuli are degraded, obscured or difficult to recognize, the amplitude of the P300 is decreased (Kok, 1985, Kok, 1980), [6].

In addition to the extensive work on describing the P300, a growing body of ERP evidence reveals ERP neural signatures of target recognition and discrimination as early as 150-200 milliseconds post-stimulus (Hopf, 2002, Vogel and Luck, 2000), [7, 8]. Johnson and Olshausen [9] demonstrated an early object recognition arising around 135 ms when low-level feature discrimination was present. These studies suggest that basic discriminative processing (e.g. differentiating faces from words, animals from non-animals, shape or color distinctions) is performed so rapidly that it must be accomplished in one feed-forward sweep of activity propagated through the visual system integrating basic visual processing with top-down template models [10]. These recent investigations quantify the difference between target and non-target ERPs to reveal distinctive ERP signatures occurring as early as 150-200 ms post-stimulus and maintained for up to 800-1000msec. post-stimulus. The differences have been identified following the presentation of objects that vary only in their target status. These target-related neural signatures provide an index of the time when object recognition is sufficiently complete for the brain to initially discriminate "targetness". These early target-related differences may reflect facilitated sensory processing (i.e., enhanced neural responses associated with matching to a top-down target template) or to decision-related post-sensory processing and recognition.

2 Method

An extensive literature review was conducted to assist in determining the appropriate classification of VE fidelity components, the trade-offs between VE fidelity components and overall VE fidelity, ways in which fidelity components can be objectively measured, and which components are most likely to have a significant impact on an observation task. This review included an investigation of human information processing (HIP) and visual perceptual skills; as well as prior research relating performance differences to various levels of VE fidelity, physiological assessment during VE-based tasks, and the effects of photorealism on task performance.

A study was then designed to determine whether physiological measures could be used to detect simulation fidelity. The experimental design and VE task environment were developed based on the literature review and resulting targeted objectives. A static, VE-based visual search task consisting of militarily-relevant vehicles in low and high fidelity conditions was developed using computer-aided drafting (CAD) software.

The stimuli consisted of a series of images containing 4 objects, one in each corner of the screen. At least 3 of the objects were distractors; the fourth was either a distractor or the target object. The target object was identified prior to the trials and remained consistent throughout the trials. In the low fidelity (LoFi) condition, minimal polygon count was used, with each object ranging from 9-14 traingles depending on its inherent complexity, and no contrast existed within each object. A sample LoFi

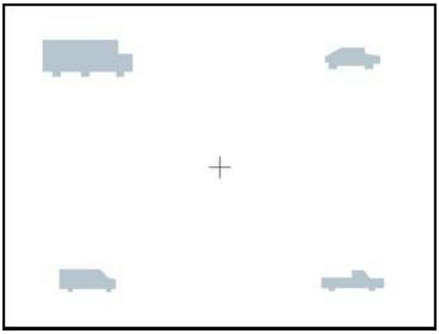


Fig. 1. Sample low fidelity (LoFi) stimulus

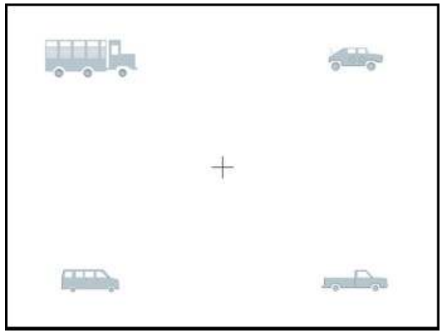


Fig. 2. Sample high fidelity (HiFi) stimulus

stimulus is shown in Figure 1. In the high fidelity condition (HiFi) the polygon count of each object was increased by 30x (+/-3%) and contrast was added, emphasizing depth and contours within each object. In both conditions the object color and background color remained constant (gray and white, respectively). A sample HiFi stimulus is shown in Figure 2.

At the start of the experiment, subjects were informed that they would be shown a series of graphics of varying detail containing a cross in the center of the screen and four objects selected from the following: battle tank, commercial truck, pick-up truck, humvee, and van. An instruction screen (shown in Figure 3) was then displayed, providing the subjects with a likeness of target objects for the low and high fidelity conditions, as well as instructions to keep their eyes focused on the cross throughout the search task.

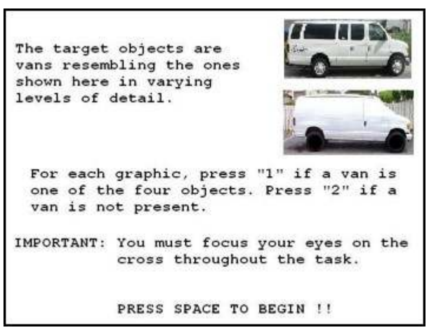


Fig. 3. Task Instruction Screen

Low and high fidelity images were then displayed in random order on a 19-inch monitor. Subjects were positioned 30 inches from the display. Approximately 50% of the presentations contained a target object.

A total of 12 participants each performed two 10-minute consecutive trials consisting of 200 stimuli presentations per trial. Each image was displayed for 2 seconds, with a 1 second inter-stimulus interval (ISI). Trials in which the participant did not provide a response within the 2-second display period were considered missed trials; these trials were reported separately from incorrect and correct response trials.

Physiological measures collected during the trials included electroencephalogram (EEG), heart rate, galvanic skin response, and eye tracking. The B-Alert® wireless Sensor Headset from Advanced Brian Monitoring (ABM) was used to acquire EEG data from 9 sites (F3, F4, C3, C4, P3, P4, Fz, Cz, and POz), referenced to linked-mastoids. The Wearable Arousal Meter (WAM) collected heart rate data, which was used to calculate arousal. Galvanic Skin Response (GSR) was assessed using the Thought Technologies Procomp System; and Eye tracking was measured via an Arrington system. A DLL was implemented to allow the EEG signal to be synchronized with the other physiological measures and the task stimuli, which were presented within a custom program using E-prime experiment management software. In addition to the task performance and physiological data, a post-task questionnaire was given to the participants.

The independent variable for this task was the fidelity condition (low or high). The dependent variables included the physiological response data, as well as the task performance data (i.e., reaction time and accuracy).

3 Results

3.1 Physiological Results

Eye Tracking. For this preliminary study, the eye-tracking data was used for the purpose of identifying if and when subjects looked away from the cross in the center of the screen during the task. Of the 12 subjects, 6 consistently looked around the screen at the stimulus objects, while the remaining 6 kept their eyes fixated on the center of the screen.

EEG. Initial data analysis was conducted for the 6 subjects that completed the task as instructed, without moving their eyes from the cross in the center of the screen, and included only the midline electrode sites (Fz, Cz and POz) as a preliminary assessment. Absolute/relative power spectral density (PSD) variables were computed for each 1-second epoch. Metrics for “engagement” and “workload” were calculated using quadratic and linear discriminant function analyses of model-selected PSD variables (1-Hz bins,1-40Hz). Event Related Potentials (ERPs) were derived based on time-locking to the presentation of the stimuli (1-second post-stimulus) or to the 1-second prior to the response. acquired from 9 scalp sites at 256 samples/sec.

ERP waveforms were combined into grand averages. All ERP waveforms were computed using only trials on which the subject correctly identified the test stimulus as either a target or a nontarget, and all were time-locked to the presentation of the test stimulus. Before averaging, all data were artifact rejected on a trial-by-trial basis for eyeblinks, excursions and excessive muscle activity using automated in-house software [11]. Trials with predominant alpha activity (present in two of the six participants) were not eliminated to allow for sufficient numbers of trials in each

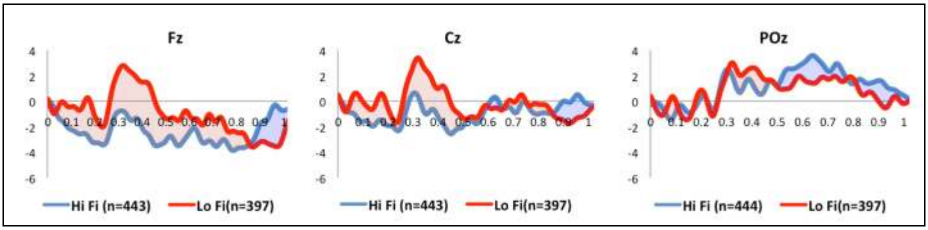


Fig. 4. Averaged ERP data for midline electrodes

average. No additional smoothing or filtering was applied on this preliminary investigational analysis. Figure 4 presents the averaged ERP waveforms for the Fz, Cz, and POz electrodes.

ERP templates for high fidelity/low fidelity (HiFi/LoFi) and targets/non-targets (T/NT) were consistently identified across the six participants. Distinctions between the HiFi/LoFi T/NT ERP templates were evident as early as 200ms post-stimulus onset and were sustained for windows in excess of 900ms post-stimulus. Maximal T/NT differences varied from 400-900msec No attempt was made in this pilot study to sort ERPs based on reaction times, which varied significantly within- and between participants (mean reaction times ranging from 550 – 1450 ms.).

Preliminary data analysis in the present study suggests that an early (onset of 200-250msec. post-stimulus) frontal-central positivity is present for all correctly identified targets and non-targets with an increase in amplitude for degraded (or low fidelity) stimuli across all 6 subjects. A much later parietal positivity (peaking between 500-700ms post-stimulus) which is likely to be a true P300 or P3b type component is evident for correct targets and is of higher amplitude for high fidelity stimuli when compared to the low fidelity. This late P300 component confirms previous reports (Kok, 1985) of degraded stimuli eliciting reduced amplitude P300.

Other Results. Arousal levels were averaged and a significant fidelity effect was found for 1 of the 6 subjects who performed the task without moving their eyes. GSR cannot be mapped to individual trials due to latency.

3.2 Performance Results

No significant effect was found as measured by reaction time and accuracy of responses for fidelity or fixation conditions; however, the task was quite simple, and thus a ceiling effect was evident.

4 Discussion

While no significant performance effects were found across conditions, consistent and detectable differences in ERP data were observed for subjects performing the visual search task in low and high fidelity conditions. Accurate identification of HiFi vs. LoFi targets was shown to elicit distinctive ERP components. Two components distinguish LoFi from HiFi: early frontal-central (250-500ms) and late parietal

(500-800ms). The early frontal-central positivity clearly distinguished the LoFi from HiFi ERPs for all participants. Though preliminary, these data suggest evidence of an early feature extraction process. Based on studies of ERPs during visual search, where individuals scan through a set of visual stimuli for a particular target, Luck and Hillyard, (1994) proposed that a spatial filtering process conducts a preliminary analysis of the stimulus array containing relevant features. They identified a component of the visual ERP in the range of 200-250msec. post-stimulus elicited by visual search arrays that varied in accordance with the filtering of distractors. When a target was distinguished by a salient feature, spatial filtering began approximately 175 ms after search onset. Second, the filtering process is dependent on the outcome of a preliminary stimulus analysis capable of rejecting non-targets on the basis of simple feature information. Alternatively the early positivity may be a reflection of working memory processes in prefrontal cortex.

The arousal data results demonstrated that some measures are more sensitive to fidelity variations than others. Some sensors can be considered as global measures, such as the skin conductance and arousal. The EEG is more specific and localized. In future studies, the eyetracking data will also be used to compare pupilometry during low and high fidelity conditions.

These results suggest that psychophysiological measures, specifically ERP, may provide a more sensitive and objective measure than traditional metrics for determining VE fidelity requirements. This research is currently being leveraged within a perceptual skills VE task in which performance is impacted significantly by fidelity degradation. Future research will compare physiological data collected in equivalent real world (RW) and VE tasks to further determine the impact of various fidelity components on task performance and training transfer.

References

1. Lathan, C.E., Tracey, M.R., Sebrechts, M.M., Clawson, D.M., Higgins, G.: Using virtual environments as training simulators: Measuring transfer. In: Stanney, K. (ed.) Handbook of Virtual Environments: Design, Implementation, and Applications, pp. 403–414. Lawrence Erlbaum Associates, Mahwah (2002)
2. Sebrechts, M.M., Lathan, C.E., Clawson, D.M., Miller, M.S., Trepagnier, C.: Transfer of training in virtual environments: issues for human performance. In: Hettinger, L.J., Haas, M.W. (eds.) Virtual and Adaptive Environments; Applications, Implications, and Human Performance Issues, pp. 67–90. Lawrence Erlbaum Associates, Mahwah (2003)
3. Waller, D., Hunt, E., Knapp, D.: The transfer of spatial knowledge in virtual environment training. Presence: Teleoperators and Virtual environments 7(2), 129–143 (1998)
4. Sutton, S., Braren, M., Zubin, J., John, E.R.: Evoked-potential correlates of stimulus uncertainty. Science 150, 1187–1188 (1965)
5. Donchin, E., Coles, M.G.H.: Is the P300 component a manifestation of context updating? Behavioral and Brain Sciences 11, 357–374 (1988)
6. Verleger, R.: Event-related potentials and cognition: A critique of the context updating hypothesis and an alternative interpretation of P3. Behavioral and Brain Sciences 11, 343–427 (1988)
7. Thorpe, S., Fize, D., Marlot, C.: Speed of processing in the human visual system. Nature 381, 520–522 (1996)

8. Fabre-Thorpe, M., Delorme, A., Marlot, C., Thorpe, S.: A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. Journal of Cognitive Neuroscience 13, 171–180 (2001)
9. Johnson, J.S., Olshausen, B.A.: Timecourse of neural signatures of object recognition. Journal of Vision, 3(7), 499-512 (2003), http://journalofvision.org/3/7/4/
10. VanRullen, R., Thorpe, S.J.: The time course of visual processing: From early perception to decision-making. Journal of Cognitive Neuroscience 13, 454–461 (2001b)
11. Berka, C., et al.: Real-time Analysis of EEG Indices of Alertness, Cognition and Memory with Wireless EEG Headset. International Journal of Human-Computer Interaction 17(2), 151–170 (2004)

Can Neurophysiologic Synchronies Provide a Platform for Adapting Team Performance?

Ronald H. Stevens[1], Trysha Galloway[1], Chris Berka[2], and Marcia Sprang[3]

[1] UCLA IMMEX Project 5601 W. Slauson Ave. #255, Culver City, CA 90230
[2] Advanced Brain Monitoring, Inc. 2237 Faraday Ave., Suite 100, Carlsbad, CA 92008
[3] Placentia-Yorba Linda Unified Sch. Dist. 1830 N. Kellogg Dr., Anaheim, CA 92807
immex_ron@hotmail.com

Abstract. We have explored using neurophysiologic patterns as an approach for developing a deeper understanding of how teams collaborate when solving time-critical, complex real-world problems. Fifteen students solved substance abuse management simulations individually, and then in teams of three while measures of mental workload (WL) and engagement (E) were generated by electroencephalography (EEG). High and low workload and engagement levels were identified at each epoch for each team member and vectors of these measures were clustered by self organizing artificial neural networks. The resulting patterns, termed neurophysiologic synchronies, differed for the five teams reflecting the teams' efficiency. When the neural synchronies were compared across the collaboration, segments were identified where different synchronies were preferentially expressed. This approach may provide an approach for monitoring the quality of team work during complex, real-world and possible one of a kind problem solving, and for adaptively modifying the teamwork flow when optimal synchronies are not frequent.

Keywords: Collaboration, EEG, Neurophysiologic synchrony.

1 Introduction

A current challenge in studying collaborative teamwork is the measurement of team cognition and the separation of it from aspects of individual cognition [16]. Research on teamwork and cooperative behaviors often adopts an input-process-output framework (IPO). In this model the interdependent acts of individuals convert inputs such as the member and task characteristics to outcomes through behavioral activities directed toward organizing teamwork to achieve collective goals. These activities are termed team processes and include goal specification, strategy formulation, systems and team monitoring, etc [15].

Much of this teamwork research has made use of externalized events focusing on *who* is a member of the team, *how* they work together and *what* they do to perform their work. The studies often rely on post-hoc elicitation of the subjective relationships among pertinent concepts. There have been fewer studies looking at the *when* of teamwork interactions although the dynamics of team function are known to be complex [4] with temporal models of teamwork suggesting that some processes transpire

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 658–667, 2009.

more frequently in action phases and others in transition periods [1-5]. Closely related to team processes are dynamic states that characterize properties of the team that vary as a function of team context, inputs, processes and outcome. Emergent states describe cognitive, motivational and affective states of teams and can serve both as outputs and inputs in dynamic IPO models. When viewed this way, the focus shifts to when and how fast activities and change occur, and the variables move from amounts, dependencies and levels to pace, cycles and synchrony [6].

One framework for studying the *when* of team cognition is macrocognition [7] which is defined as the externalized and internalized high-level mental processes employed by teams to create new knowledge during complex collaborative problem solving. External processes (processes occurring outside the head) are those associated with actions that are observable and measurable in a consistent, reliable, repeatable manner. Internalized processes are those that cannot be expressed externally and are generally approached indirectly through qualitative metrics like think aloud protocols or surrogate quantitative metrics, (pupil size, EEG metrics, galvanic skin responses). To our knowledge, there have been no reports linking neurophysiologic correlates of internalized processes across members of a team as they engage in teamwork tasks. This however would seem to be an important contribution to the goal of better understanding the construct of team cognition.

Our hypotheses is that as members of a team perform a collaborative task each will exhibit varying degrees of cognitive components such as attention, workload, engagement, etc. and the levels of these components at any one time will depend (at least) on 1) what that person was doing at a particular time, 2) the progress the team has made toward the task goal, and 3) the composition and experience of the team. Given the temporal model of team processes, some of the balances of the components across team members may also repeat as different phases of the task, like data acquisition, or communication are repeatedly executed. In this study we have directly tested these hypotheses using EEG measures of mental workload and engagement.

2 Tasks and Methods

2.1 IMMEX Substance Abuse Simulations (*SOS*)

The collaboration task is an IMMEX™ problem set called *SOS* which are a series of substance abuse simulations cast in a reality show format [8-10]. The case begins with a short introduction to a person who may / may not be abusing drugs. The challenge for the student is to gather sufficient information about this person to answer the question "Should this person seek help, and if so, from whom?" The primary interface is a timeline that covers up to twelve specific events (such as health, job, social school, etc. related activities) and drilling down into this interface provides information in eleven areas with contents covering subject history, behavior, medical data and conjecture, and help. These 600+ content items are divided into social and scientific areas allowing the student to gather information from many perspectives. Prior modeling studies have shown that ~20% of the students use science-only approaches, ~40% will use social approaches, and ~40% will use a combination of the two. This task provides a convenient mechanism for the division of teamwork (i.e. social vs. scientific evidence), as well as a potential source of conflict within the group as to what evidence is important relative to the decision.

Experimentally, students log on to IMMEX™ and individually perform a *SOS* simulation so that each can develop a mental model of the problem space, and so that individual levels of EEG-related workload and engagement can be determined. Two students then log on to a second *SOS* problem set where Member A selects data from the timeline and reports information from General Health, Anecdotes and Cell & e-mails (i.e. the social perspective), Member C selects data from all the other science categories and reports them to the group (the science perspective) and the leader (Member B) integrates the information and decides when to make a decision, and what the decision will be. The time allowed is 30 minutes (a time constraint).

2.2 EEG Metrics

The EEG data acquired from the wireless headset developed by Advanced Brain Monitoring, Inc. uses an integrated hardware and software solution for acquisition and real-time analysis of the EEG. It has demonstrated feasibility for acquiring high quality EEG in real-world environments including workplace, classroom and military operational settings. The system contains an easily-applied wireless EEG system that includes intelligent software designed to identify and eliminate multiple sources of biological and environmental contamination and allow real-time classification of cognitive state changes even in challenging environments. The 9-channel wireless headset includes sensor site locations: F3, F4, C3, C4, P3, P4, Fz, Cz, POz in a monopolar configuration referenced to linked mastoids. ABM B-Alert® software acquires the data and quantifies alertness, engagement and mental workload in real-time using linear and quadratic discriminant function analyses (DFA) with model-selected PSD variables in each of the 1-hz bins from 1-40hz, ratios of power bins, event-related power (PERP) and/or wavelet transform calculations.

To monitor "mental workload" (WL) and "engagement" (E) using the B-Alert model EEG metrics, values ranging from 0.1-1.0, are calculated for each 1-second epoch of EEG. Simple baseline tasks are used to fit the EEG classification algorithms to the individual so that the cognitive state models can then be applied to increasingly complex task environments, providing a highly sensitive and specific technique for identifying an individual's neural signatures of cognition in both real-time and offline analysis. These methods have proven valid in EEG quantification of drowsiness-alertness during driving simulation, simple and complex cognitive tasks and in military, industrial and educational simulation environments, quantifying mental workload in military simulation environments, distinguishing spatial and verbal processing in simple and complex tasks, characterizing alertness and memory deficits in patients with obstructive sleep apnea, and identifying individual differences in susceptibility to the effects of sleep deprivation [11-13].

2.3 Experimental Protocol

The data flow (Figure 1) is organized into Collection, Processing, Modeling and Analysis modules. The teams perform the *SOS* collaborative tasks while EEG is being collected at 256 Hz from 6-electrode portable headsets. The data Collection initiates

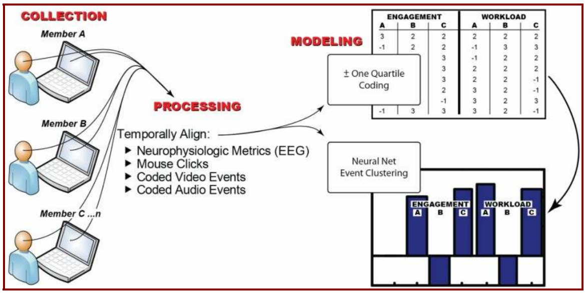


Fig. 1. Outline of Experimental Protocol

with the start of the *SOS* simulations on the time synchronized computers of the two team members. The computers also run Morae (Techsmith, Inc.) which records a video and audio trace of each participant and generates logs with timestamps of mouse clicks screen refreshes, etc. The Processing module aligns the EEG logs containing the second-by-second WL and E values from each of the three team members and interleaves them with mouse clicks logs and video/audio logs.

The values of WL and E were determined for the individual performances of each student, as well as for each student during the collaboration event. As shown in Figure 2, IMMEX tasks are complex eliciting more WL from the students than on a 3-choice vigilance task (3-CVT) baseline task. The subjects also expend more WL in a teamwork situation than they did when performing the task individually, which may relate to the process cost of collaboration discussed by others [16].

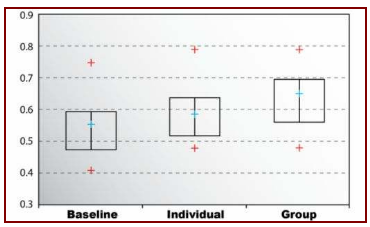


Fig. 2. EEG-WL Levels During Baseline, Individual and Group Conditions. The levels of WL were calculated for 15 individuals on a 3-CVT task, during an individual performance of an SOS problem, and during a 3-person team performance.

The values of WL and E were then normalized for each team member by statistically partitioning them into the upper quartile, the lower quartile, and the half in the middle representing high, low and average levels of WL and E. These partitions were assigned the values 3, -1, and 2 and were combined for each of the members of the team to create training vectors (Figure 1, Modeling) for training self organizing artificial neural networks (ANN) as previously described [8,9]. This process results in patterns of WL and E measures across the members of the team on a second by second time scale. We define these epochs of alignment as neurophysiologic synchronies.

3 Experimental Results

3.1 Team Differences in Neurophysiologic Patterns of Collaboration

We first examined the performances of five collaboration groups to identify common and dissimilar neural synchronies (i.e. combinations of WL and E across team members) across teams. An example of this analysis is shown in Figure 3 where an ANN was trained with the neural synchronies from 5 different groups. The output from such an analysis is a series of ANN nodes each representing a synchrony with a different profile of neurophysiologic indicators. After training, twenty three of the twenty five nodes contained between 37 and 562 epochs with different patterns of neurophysiologic synchrony of WL and E. The most common synchrony was represented by nodes 14 and 8 which consisted of epochs where all members were engaged and working at moderate to high levels. This may represent the nature of the IMMEX task itself which requires more workload than simpler image identification tasks [18]. Other frequent synchronies were nodes 23, 4, and 2 where one of the members was either not working hard or not highly engaged.

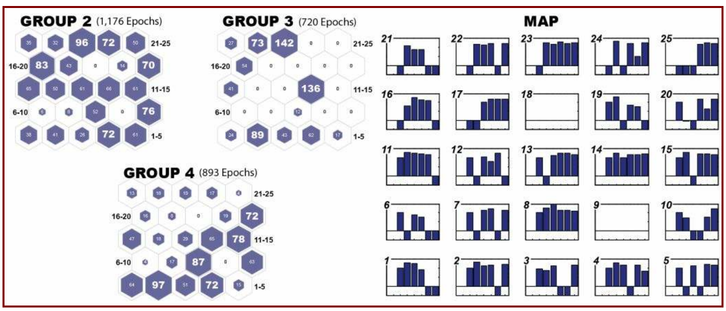


Fig. 3. Neural Synchrony Patterns across Teams. A self organizing ANN was trained with the collaboration performances of 5 teams and retested with the individual performances. The numbers in the hexagons reflect the number of times the pattern was repeated during the task.

When the different teams were tested on this combined ANN they showed significant differences in the proportions of neural synchronies being expressed. Group 3 for instance showed a pattern of synchronies restricted to only half of the neural network nodes. Many of the epochs reflected times where the whole team was engaged or working, or where only Team Member A was minimally engaged (i.e. node 23). Group 4 in contrast showed a greater diversity of neurophysiologic synchronies. There were few epochs clustered at node 23 and instead showed more epochs at nodes 1 and 2 where the common feature was low engagement of the Team Leader, and nodes 10, 15 and 20 where Team Member B was not engaged. Group 2 was more diverse still showing similarities with both Group 3 (i.e. node 23) and Group 4 (i.e. nodes 4, 10 and 13).

3.2 Do Common Neurophysiologic Patterns Have Collaborative Significance?

During collaboration effective teams execute processes that often occur in a cyclical fashion depending on task demands. In a second set of studies we tried to determine if the different patterns of WL and E expression across the team had significance vis a vis the collaboration event. Most team tasks, including the IMMEX problem solving tasks, can be separated into segments consisting of mental model formation, mental model sharing and integration, and mental model consensus and revision. These can be further divided into behavioral episodes relating to team processes. Figure 4 shows the task breakdown for one collaborative team (Group 2) (1178 epochs or seconds duration). The tasks included the reading of the task and initial discussions, explorations of the problem space, deriving a consensus regarding the decision, etc. We have highlighted these tasks by the different stages of mental model formation, sharing and integration, and convergence and revision. The epochs reflecting different team synchronies were temporally aligned with the collaborative events. The most common synchrony (113 epochs) showed limited mouse click activity, all three members were experiencing elevated WL and the Team Leader and Member C were highly engaged.

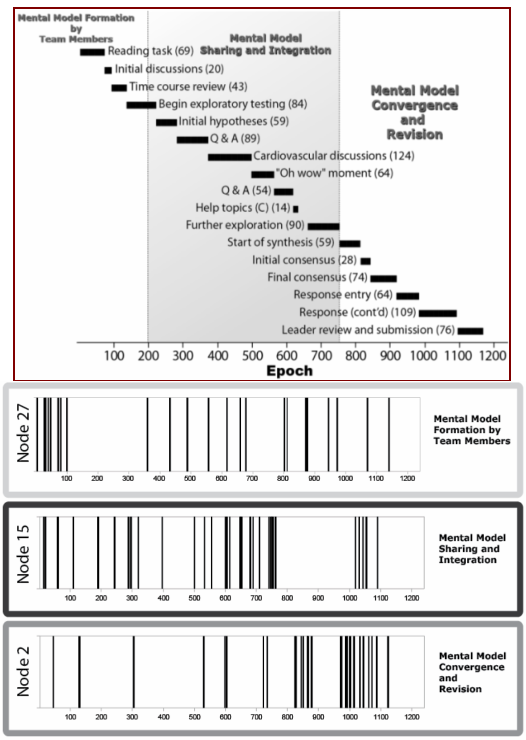


Fig. 4. (Top) Team Behaviors during a Sample *SOS* Collaboration Session. The numbers in parentheses indicate the number of epochs for each task. (Bottom) Selective expression of neurophysiologic synchrony patterns during different segments of the collaboration task.

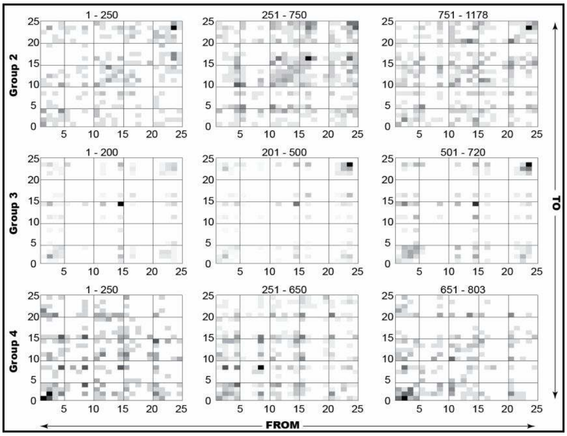


Fig. 5. Temporal Analysis of Nodal Transitions. A nearest neighbor correlation analysis was performed for the beginning, middle and end of the collaborations for three groups. The diagrams show the transitions from one nodal pattern (X-axis) to another (Y-axis).

This profile was present throughout the collaborative task and may reflect a common feature of this team's interaction. In this regard, examination of the video log indicates that interactions between the leader and team member A were less frequent than interactions with team member C. Neurophysiologic synchronies identified by other neural network nodes were more selectively expressed during the task with some being preferentially expressed during the mental model forming stage whereas others were more prevalent during the mental model convergence and revision stage (chi square = 1291, p=< 0.001).

A second approach examined the autocorrelations of the synchronies with a time lag of 1, i.e. a sequential nearest neighbor analysis asking 'If a synchrony pattern is being expressed, what pattern is likely to follow next?' The diagrams in Figure 5 are called From >To diagrams and indicate the transition from a node on the X axis to a node on the Y axis. Similarly, to determine how a node was arrived at, a Y value can be traced across the X axis. Figure 5 shows such an analysis for Groups 2, 3 and 4. To relate the correlations to different stages of the collaborative task, the analyses were repeated for the early, middle and late epochs of the teamwork as indicated by the epoch numbers above each diagram. Group 2 showed the lowest From-> To correlations (-.14, .19 and -.25 for the early, middle and late epochs), had the lowest proportion (12%) of synchronies where all members were simultaneously engaged and working (i.e. nodes 8 and 14), and also took the longest to complete the task. The most frequent patterns were where the E of Team Member A was low while the other members were fully engaged and working. Group 3 showed the most restricted pattern of synchrony, had the highest From->To correlations and the highest proportion of synchronies (19%) where all members were fully engaged and working. The transition from Node 14 to 14 dominated early during the collaboration, and transited to a Node 23->23 transition indicating a state where the engagement of Team Member A was reduced while the others were engaged and working. The autocorrelations were .77, .58 and .75 respectively for the early, middle and late epochs of the collaboration.

In this group Team Member A also had the lowest overall WL and E levels and ordered fewer items during the simulation than did the second team member (186 vs. 234 tests ordered). Of the five teams tested, Group 3 was the most effective as judged subjectively from the video logs, as well as objectively with the most rapid solution time (11 minutes), and the final answer.

Group 4 displayed an intermediate diversity of neurophysiologic synchronies compared with the other groups and this was also reflected in the From->To correlations. The dominant nodal patterns were nodes 1, 2 and 8, where node 8 is similar to node 14 with all members are engaged and working, and this constituted 15% of the total number of epochs. During the initial part of the teamwork the time-lagged correlation was .22 indicating a less stable pattern than for Group 3. The major nodal transitions were from nodes 1 to 1 and nodes 2 to 2, and the common feature of nodes 1 and 2 is the decreased WL levels in the Team Leader. During the middle portion of the teamwork the nodal correlations increased to .59 with the dominant repeating nodes being 8 and 4. During task closure the timed lagged correlation dropped to .44 the repeating node 8 transition decreased and the transition from node 2 to node 1 returned. In Group 4, the Team Leader had the lowest overall WL of any of the team members and the second highest E levels.

4 Discussion

This study describes our preliminary efforts at determining if neurophysiologic synchronies can be observed during problem solving teamwork. We define neurophysiologic synchronies as the coordinated expression of different levels of neurophysiologic indicators by individuals of a team as they engage in collaborative activities. In this study we have used the neurophysiologic correlates of workload and engagement as defined by the B-Alert EEG system, although there is no a priori reason that other measures could not be used, or included. The studies to date, while involving only five teams, suggest that patterns of neurophysiologic synchrony can be observed in different teams which may have collaborative significance. An important next step is to link them to other collaboration behaviors, and an important challenge will be determining the granularity to conduct these studies. The enrichment of some patterns at the early and late stages of the teamwork suggests a temporally related contribution which may relate to different aspects of the collaboration task. A more granular approach would be to link the synchronies to common behaviors in IMMEX™ such as the ordering of tests by mouse clicking on menu items or other behaviors such as questioning, responding, etc. Such epoch "tagging" may facilitate categorizing the macrocognitive constructs that are occurring simultaneously such as synthesis, questioning, team consensus, revision / analysis, etc. Neurophysiologic synchronies may also be useful for adaptively establishing or modifying the balance of team members and their degree of participation. Situations where a member is consistently lower in WL and/or E while the other members are fully engaged and working hard may indicate a less effective team member. This may be particularly important as the efficiency of a team is in completing a task (as measured by time to completion) was proportional to the percentage of neural synchronies where all members are both engaged and working. (i.e. nodes 8 and 14). Another possible indicator of effective / ineffective teams may be the persistence of neural synchronization indicated by the degree of correlations between synchronies with a time lag of 1 epoch. These nearest neighbor correlations may indicate that a team state is more stable over a longer period of time, while teams with low or negative From -> To correlations may represent teams where the members are searching for an effective rhythm. The nodal neurophysiologic From -> To correlations may also make this approach amenable to the development of dynamic and predictive models either through Hidden Markov Modeling [9] or through a more dynamical systems approach such as phase space reconstruction [14].

Finally, the studies may also provide a tool for approaching the process cost associated with teamwork. Team workload is a core component of most theories of collaborative and cooperative learning, and is described as the resources available by a team for a task relative to the demands placed on it. As with individuals, team performance is presumed to deteriorate when the task demands exceed available resources. Experimental evidence suggests that this may be so, with the higher the workload of the least-loaded team member, the lower the team performance [17]. Many factors can contribute to the workload of a member of a team and the overall team functioning. At one extreme, the individual may have difficulty with his own task which would lead to individual task overload. Depending on the degree of critical nature of that task for the overall team goal, this may or may not have an effect on team outcome. At the other pole, there may be disruptions in the degree of information sharing leading to negative team performance.

Workload in teams, however, is complex and at its simplest consists of the workload of a team member on his/her individual task within the team (Task Awareness) as well as more of a team process workload (Teamwork Awareness) which relates to the resources required to be an active member of a team. While the ideas of workload and work overload are practically appealing, it has been difficult to derive quantitative measures of them. The results in Figure 2, suggest that the EEG-WL metric may provide a useful measure for this added cost.

References

1. Cannon-Bowers, J.A., Salas, E., Converse, S.A.: Shared mental models in expert team decision making. In: Castellan, J.N.J. (ed.) Current issues in individual and group decision making, pp. 221–246. Lawerence Erlbaum, Hillsdale (1993)
2. Cohen, S.G., Bailey, D.E.: What makes teams work: Group effectiveness research from the shop floor to the executive suite. Journal of Management 23, 239–290 (1997)
3. Cooke, N.J., Kiekel, P.A., Salas, E., Stout, R.: Measuring team knowledge: A window to the cognitive underpinnings of team performance. Journal of Applied Psychology 7, 179–199 (2003)
4. Mathieu, J., Maynard, M.T., Rapp, T., Gilson, L.: Team Effectiveness 1997-2007: A Review of Recent Advancements and a Glimpse Into the Future. Journal of Management 34(3), 410–476 (2008)
5. Mohammed, S., Klimoski, R., Rentsch, J.: The Measurement of Team Models: We Have No Shared Scheme. Organizational Research Methods 3, 123–165 (2000)

6. Ancona, D.G., Chong, C.L.: Cycles and synchrony: The temporal role of context in team behavior. In: Wageman, R. (ed.) Groups in context, pp. 33–48. JAI Press, Stamford (1999)
7. Warner, N., Letsky, M., Cowen, M.: Cognitive Model of Team Collaboration: Macro-Cognitive Focus. In: Proceedings of the 49th Human Factors and Ergonomics Society Annual Meeting, Orlando, FL, September 26-30 (2005)
8. Stevens, R.H.: A Value-Based Approach for Quantifying Student's Scientific Problem Solving Efficiency and Effectiveness Within and Across Educational Systems. In: Lissitz, R.W. (ed.) Assessing and Modeling Cognitive Development in School, pp. 217–240. JAM Press, Maple Grove (2007)
9. Stevens, R.H., Soller, A., Cooper, M., Sprang, M.: Modeling the Development of Problem Solving Skills in Chemistry with a Web-Based Tutor. In: Lester, J.C., Vicari, R.M., Paraguaca, F. (eds.) 7th International Conference Proceedings Intelligent Tutoring Systems, pp. 580–591. Springer, Heidelberg (2004)
10. Stevens, R.H., Casillas, A.: Artificial Neural Networks. In: Mislevy, R.E., Williamson, D.M., Bejar, I. (eds.) Automated Scoring of Complex Tasks in Computer Based Testing: An Introduction, pp. 259–312. Lawrence Erlbaum, Mahwah (2006)
11. Levendowski, D.J., Westbrook, P., Berka, C., et al.: Event-related potentials during a psychomotor vigilance task in sleep apnea patients and healthy subjects. Sleep 25, A462–A463 (2002)
12. Stevens, R.H., Galloway, T., Berka, C.: Integrating Innovative Neuro-educational Technologies (I-Net) into K-12 Science Classrooms. In: Schmorrow, D.D., Reeves, L.M. (eds.) HCII 2007 and FAC 2007. LNCS, vol. 4565, pp. 47–56. Springer, Heidelberg (2007)
13. Berka, C., Levendowski, D.J., Ramsey, C.K., Davis, G., Lumicao, M.N., Stanney, K., Reeves, L., Regli, S., Tremoulet, P.D., Stibler, K.: Evaluation of an EEG-Workload Model in an Aegis Simulation. In: Proceedings of the SPIE Defense and Security Symposium, Biomonitoring for Physiological and Cognitive Performance during Military Operations 2005, vol. 5797, pp. 90–99 (2005)
14. Abarbanel, H.D.: Analysis of Observed Chaotic Data. Springer, New York (1995)
15. Marks, M.A., Mathieu, J.E., Zaccaro, S.J.: A Temporally Based Framework and Taxonomy of Team Process. The Academy of Management Review 26(3), 356–376 (2001)
16. Cooke, N.J., Gorman, J.C., Preston, A.K.: Communication as Team-level Cognitive Processing. In: Letsky, M.P., Warner, N.W., Fiore, S.M., Smith, C.A.P. (eds.) Macrocognition in Teams, pp. 51–64. Ashgate Publishing, Burlington (2008)
17. Urban, J.M., Bowers, C.A., Monday, S.D., Morgan Jr., B.B.: Workload, team structure, and communication in team performance. Military Psychology 7(2), 123–139 (1995)
18. Stevens, R.H., Galloway, T.L., Berka, C.: Allocation of time, EEG-Engagement and EEG-Workload as scientific problem solving skills are acquired. In: Proceeding Human Factors and Ergonomics Society, 51st Annual Meeting, Baltimore, MD, October 1-5 (2007)

Seeing the World through an Expert's Eyes: Context-Aware Display as a Training Companion

Marc T. Tomlinson, Michael Howe, and Bradley C. Love

Department of Psychology, University of Texas at Austin, Austin Texas
{marctomlinson,Michael.howe,Brad_love}@mail.utexas.edu

Abstract. Responsive Adaptive Display Anticipates Requests (RADAR) is a domain general system that learns to highlight an individual's preferred information displays, given the current context. Previous studies with human subjects in a video game environment demonstrate that RADAR is an effective cognitive aid. RADAR increases situation awareness and reduces cognitive load by anticipating and providing task relevant information. Additionally, because RADAR's fit to a user's behavior encapsulates the user's situation-driven information preferences, RADAR also excels as a descriptive and predictive assessment tool. Here, we focus RADAR as a training aid. We test the hypothesis that novices can benefit from training under a RADAR model derived from an expert's behavioral patterns. The results indicate that novices exposed to an expert's information preferences through RADAR rapidly learn to conform to the expert's preferences.

1 Introduction

When boarding an airplane, a furtive look into the cockpit reveals a vast array of dials, displays, and controls. The expert pilot can make sense of this array of options and can appreciate when each instrument is relevant to operating the aircraft. For example, expert pilots know which gauges are relevant to different phases of flight. In this article, we discuss a context-aware approach to information display named Responsive Adaptive Display Anticipates Requests (RADAR). RADAR learns to highlight the situation-relevant information by observing the user.

We discuss how RADAR can be used to analyze and describe individual differences in information needs, as well as present evidence that RADAR can be used to allow novices to see the world through the eyes of an expert. When training under an expert's RADAR model, we find that novices' information use patterns converge to those of the expert from whom the model was derived.

Related work has attempted to predict user information needs by correctly attributing intentions, beliefs, and goals to the user. Plan recognition models tend to subscribe to the Belief-Desires-Intention framework [1]. This line of work relies on knowledge-based approaches for user modeling and encoding insights from domain-specific experts [2]. These approaches can involve identifying a user's subgoals through task-analysis [3]. Once a user's beliefs, intentions, and goals are understood, a display can be adapted appropriately [2].

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 668–677, 2009.

Alternatively, instead of focusing on identifying the internal state of the user, some approaches rely on input from domain experts. For example, human experts can label episodes that can serve as training instances for machine learning models that prioritize display elements [4]. Alternatively, input from human experts can be used to build expert systems or Bayesian models to prioritize displays [5]. This approach relies on extensive input from human experts, and the ability of those experts to introspect on the reasons for their performance.

Our approach diverges from the aforementioned work. Rather than prescribe which information source a user should prioritize, RADAR highlights the information a user would select if the user searched through all possible options. This approach may be preferable in domains where it is unclear what is normative. Unlike work in plan recognition, we sidestep the problem of ascribing and ascertaining the user's internal mental state. Instead, RADAR learns to directly predict a user's desired display from contextual (i.e., situational) features (see Figure 1).

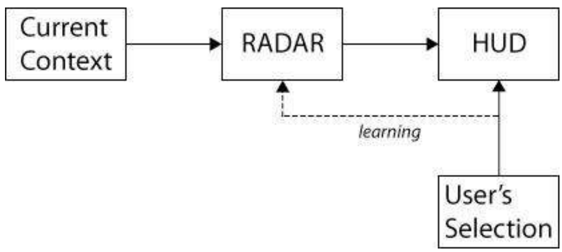


Fig. 1. RADAR takes as input the current context (e.g. Recent game history) and outputs its preferred display to the HUD. The user (e.g., the game player) can override RADAR's choice. Such corrections serve as learning signals to RADAR and increase the likelihood that RADAR will select the user's preferred display in similar situations in the future. Over time, RADAR approximates the information preferences of a specific user, allowing the user to offload the task of selecting the relevant information source (i.e. display) from numerous options.

Furthermore, RADAR emphasizes the benefits of continuous learning by the display, as opposed to preprogrammed interfaces [6]. Adopting a learning approach to an adaptive display has a number of positive consequences, including the ability to take into account individual differences across users [7]. Another positive consequence is that minimal input from subject matter experts is required to build a system. Like other context-aware applications that adopt a keyhole approach [8,9], our approach infers a user's preferences without interfering with or directly querying the user [10]. Interfaces that highlight recently selected menu items follow a similar logic [11], though our approach is more open ended in terms of possible predictors and learnable relationships from predictors to display preferences.

Whereas previous work with RADAR [12], which we review below, has evaluated RADAR as a cognitive aid and assessment tool, the current experiment evaluates RADAR's promise as a training companion. The current experiment asks whether RADAR can speed the novice to expert transition by exposing novices to the display preferences of an expert (i.e., train under an expert's RADAR model).

Our approach has some potential benefits. In some domains, knowledge can be directly elicited from experts and simple instruction can boost novice performance to expert levels [13]. However, in many domains, an expert's knowledge is not accessible by self-report [14,15]. In practice, training methods for novices that rely on both direct instruction and pattern recognition methods work best [16]. Indeed, having novices view expert solutions is more effective than even providing corrective feedback [17,18]. These findings suggest that there is more to expertise than what an expert can report verbally from introspection. This conclusion is not surprising given that human learning is subserved by multiple learning systems, only some of which are accessible to introspection and verbal report [19].

Our training goal is to make novices conform to the information preferences of experts in order to improve task performance. In service of this goal, standard verbal instructions, coupled with RADAR, provide users with training opportunities that can engage both verbal and non-verbal learning systems. A RADAR system trained based on an expert's performance data is a potentially powerful training tool for novices. Such a tool might allow a novice to become sensitive to the information preferences of an expert while performing the relevant task. Importantly, such a training system does not require eliciting explicit knowledge from an expert and can impart expert knowledge that is not readily verbalized. We present an experiment that investigates how novices trained with an expert RADAR system perform compared to those trained under a control model.

1.1 RADAR's Operation

RADAR is designed to operate in task environments in which the user must select which display among numerous displays to monitor. For example, we evaluate RADAR in an arcade game environment in which players select which of eight possible displays to show on a Head-Up Display (HUD). RADAR takes as input the current context (e.g., recent game history) encoded as a feature vector and outputs to the HUD the display it predicts the user wishes to view (See Figure 1). The user is free to override RADAR's choice. RADAR learns from the user's acceptance or rejection of its display choices and over time converges to selecting the displays the user desires. Alternatively, RADAR can observe and learn to mimic a user's display preferences offline.

RADAR employs a two-stage stochastic decision process at every time step. In the first stage, RADAR estimates the probability that a user will update the HUD given the current context. When the sampled probability from the first stage results in a display update, RADAR proceeds to the second stage (otherwise the current display remains unchanged). In the second stage, RADAR estimates the probability distribution for the next display choice given the current context, and samples this probability distribution to select the next display.

The motivation for the two-stage approach is both computational and psychological. Separating display prediction into two stages improves RADAR's ability to predict display transitions. The same display currently desired is highly likely to be desired in 250 ms. This constancy would dominate learning if both stages were combined. The second stage's focus on display transitions allows for improved estimation of these relatively rare, but critical, events.

Psychologically, the first stage corresponds to identifying key events in a continuous (unsegmented) environment, whereas the second stage corresponds to predicting event transitions. To make an analogy to speech perception, people segment the continuous speech stream into words (akin to RADAR's first stage) in the absence of reliable acoustical gaps between words [20]. Akin to RADAR's second stage, people anticipate which word (i.e., event) is likely to follow given the preceding words [21].

One view is that event segmentation serves an adaptive function by integrating information over the recent past to improve predictions about the near future (see [22], for a review). In support of this view, individuals who are better able to segment ongoing activity into events display enhanced memory [23]. People's judgments of event boundaries are reliable [24] and tend to show high agreement with others [25]. For example, two people watching a person make a peanut butter and jelly sandwich will tend to agree on the steps involved. These two people will also both segment off surprising or unexpected events, like the sandwich maker dropping the sandwich on the floor.

The probability distributions associated with both stages (event segmentation and event prediction) are estimated by simple buffer networks [26]. Buffer networks represent time spatially as a series of slots, each containing the context (e.g., game situation) at a recent time slice, encoded as a feature vector. The buffer allows both ongoing events and events from the recent past to influence display prediction. Despite their simplicity, buffer networks have been shown to account for a surprising number of findings in human sequential learning [27]. At each time step, weights from the buffer are increased from activated features to the display option shown in the HUD, whereas weights to the other display options are decreased. Over time, this simple error correction learning process approximates a user's information preferences. RADAR's weights can be used to assess individual differences and user performance. Details of RADAR's implementation are discussed elsewhere [12].

1.2 Previous Work

Previous experiments with RADAR have shown that it is an effective cognitive aid [12]. RADAR model trained from the aggregated data of several domain experts have been shown to be better at highlighting important information, than control models which only display information using the same base rates as the experts. Furthermore, when users are assisted in making display choices by an individually tailored RADAR model, their performance is better than when they are solely responsible for controlling the display.

RADAR has also demonstrated its usefulness as an assessment tool. By comparing model fits between expert and novice players, RADAR revels that there are significant differences in the pattern of information usage between the two groups. Furthermore, a novice player's success in the game is predicted by how well an expert's RADAR model fits their display choices.

As previously discussed, RADAR's first stage is hypothesized to be akin to scene segmentation. The first stage of the model learns to predict when a user chooses to update the display. The first stage is independent of the second stage which chooses the successor display. As discussed above, cognitive load and change in the environment are greatest at event boundaries (the very times one would want RADAR to update the display). Results from subjects playing our video game without RADAR support the notion that RADAR's first stage is akin to event segmentation. For an example expert subject, Figure 2 shows the mean number of feature changes in the environment over a ten second window before and after display channel changes. Figure 2 suggest that change is greatest at display updates, as they are at event boundaries. This consequence of people's interactions with the environment may explain why RADAR is effective as a cognitive aid. Interestingly, there is a lag between the change in features and the actual time of the channel change. We believe this lag arises because people are slow to respond to the changing event due to concurrent demands in the video game task.

While experts show individual differences in which channels they choose to view at any given moment, they have a remarkable level of agreement on when they should change display channels. This is reflected by assessing the fit of models created under one individual to the actual data provided by another individual. We see that the first stage of a given expert fits all experts almost as well as it fits the individual that created the model. In contrast, the second stage shows a marked decrease in fit to other experts compared to the individual's own data. Individual differences arise in information preferences (stage 2), but not in event segmentation (stage 1).

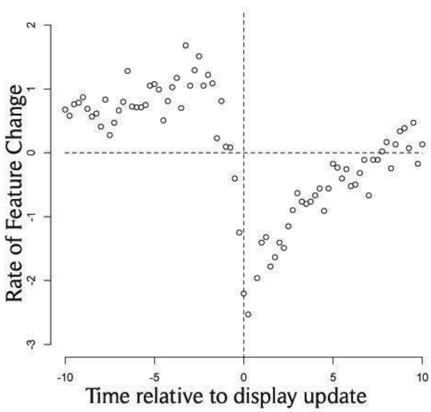


Fig. 2. Feature change (a proxy for change in the environment) is plotted in z-scores.Time on the horizontal axis (in seconds) is relative to display updates (negative is prior to update, positive is post update). The plot indicates that feature change is greatest prior to a display change. These results support the notion that display updates are akin to event boundaries.

2 Training Novices with Expert Displays

We explore the possibility that novices can learn to sample the environment like experts following training under an expert's RADAR model. An expert's RADAR model captures the expert's situational information preferences. Thus, a novice training under an expert's model can potentially benefit from the expert's perspective. Potential advantages of this approach include exposing novices to expert knowledge that is not readily verbalizable and providing expert insight in the context of performing the relevant task. To test this hypothesis, we had novices play in the tank environment with display choices determined by either an expert or control RADAR model. Subjects alternated between having displays provided to them (either by an expert or control model) and choosing displays manually. We compare manual information selections for these two groups of novice subjects.

2.1 Methods

RADAR Training Models

Subjects trained under various RADAR models. These models were built from fitting three subjects from a previous study. In the previous study, subjects played for 11 hours, controlling the display manually for the entire period. We created an expert model for each of three subjects based on the last three hours of play. Rather than use all the features available in the game, we determined the features that subjects actually entertained. This was done by evaluating subsets of all possible features using cross validation [28]. In cross validation, including features that are not psychologically real decreases performance on the data held out to test for generalization. These fits provided our three *expert* models. We then created a set of control models. The control models were specified to choose the channel that its corresponding expert model is least likely to choose. The control models also change the channel when the expert model is least likely to change, but importantly maintains the same rate of changes over time. The first stage of the control models were also decoupled from the environment, so that channel changes would not be indicative of the underlying event structure.

Design and Procedure

Thirty students were recruited from the University of Texas at Austin and were paid and given class credit for participation. The subjects played in the tank environment for three 1.5 hour sessions over a one-week period. Subjects were randomly assigned to either the expert or control condition. Subjects in the expert condition were randomly assigned to train under one of the three expert models, whereas subjects in the control condition were randomly assigned to one of the three control models. Participants in both conditions alternated between five-minute blocks of manually controlling the display and having their RADAR control the display. Which RADAR model controlled the display is the only difference in procedure across subjects.

2.2 Results

Fit of Subjects' Manual Play Data by their RADAR Model used in Training

One question is whether subjects conform to the RADAR model that they trained under. Here, we assess the probability that a subject's RADAR model correctly predicts the subject's display choices under manual play. Expert condition subjects' display choices were more accurately predicted (.25 vs. .13) by their RADAR model than were control subjects, $\mathbf{F}(1,18) = 36.10$, $\mathbf{p} < .001$. There was also a main effect of session (i.e., improvement over time), $\mathbf{F}(1,18)=4.61$, $\mathbf{p}<.05$. Importantly, there was an interaction of these two factors, such that expert condition subjects came to conform more to their model over session, $\mathbf{F}(1,18) = 8.70$, $\mathbf{p} < .01$. The left panel of Figure 3 shows that the interaction is driven by gains made by subjects in the expert condition.

Evaluating Novices' Progression Toward Expert-Like Performance

The previous analysis demonstrates that subjects come to conform to their display model, particularly subjects in the expert condition. One key question is the degree to which people come to behave like experts. To answer this question, we used all three expert models to predict each subject's display choices and averaged the fit of the three models to get a measure of expert-like behavior. There is a main effect of convergence over session, **F**(1,18) = 13.20, **p** < .01, although there is no main effect for training condition. Importantly, there is an interaction such that subjects in the expert condition become more expert-like over sessions than do subjects in the control condition, **F**(1,18) = 4.96, **p** < .05. In fact, subjects in the control condition show no significant difference in expert fit (.24 vs. .25) between the first and last session, **t** < 1, whereas subjects in the expert condition improve (.23 vs. .27) significantly, **t**(15)=3.20, **p** < .01. These results suggest that subjects in the expert condition become more expert-like in their information selections, whereas subjects in the control condition did not. Mere experience on task does not appear to guarantee the emergence of expert-like behavior in terms of display choice.

Display Updating as a Function of Training-Mode

The previous analysis focused on display choice, RADAR's second stage. One question is whether differences between expert and novice condition subjects exist in display update (i.e., when to change the display), RADAR's first stage. Analyses indicate a main effect for converging to the average expert fit over session, **F**(1,18) = 13.01, **p** < .01, but no effects of training condition were observed. The change patterns of subjects in both the expert and control conditions were fit equally well by the expert models (see the right panel of Figure 3). This result is highly suggestive that event segmentation, in contrast to display choice, is something that is learned by experience on task and is not facilitated by training under an expert model. The lack of an interaction between session and training condition also agrees with previous work that finds that different expert models' fist stages have higher inter-agreement than do their the second stages.

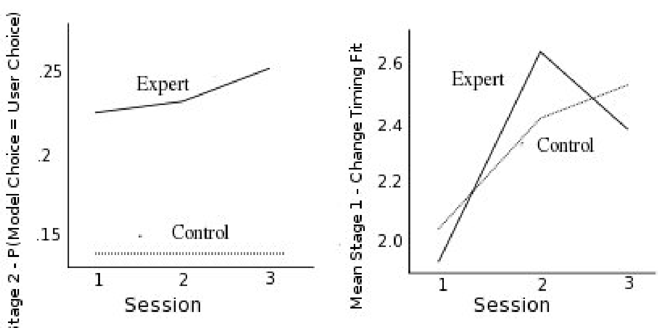


Fig. 3. The left panel shows that subjects' RADAR models better predict their display choices in the expert condition and this advantage grows with training. The right panel shows that expert and control condition subjects make display updates in roughly the same fashion. The average fit of the three expert models yields similar results for both conditions with regards to display update.

3 Discussion

Advances in information technology make large quantities of information available to human decision makers. In this deluge of information, finding and selecting the relevant piece of information imposes a burden on the user. This burden is particularly onerous for novices within complex, dynamic environments. RADAR is a domain-general system that learns to approximate the information search processes of an individual user.

RADAR contains two stages. The first stage is akin to event segmentation and determines when to update the display. The second stage determines, given a display update, which display to select. Previous work demonstrates that RADAR improves user performance [12]. Here, we report results that indicate that subjects who train under an expert's RADAR model learn to choose displays consistent with the second stage of expert RADAR models.

The same result did not hold for display update, embodied in RADAR's first stage. In the case of display update, subjects trained under expert and control RADAR models both converged to expert-like updates over time. This result supports previous research [12, 24, 25] demonstrating reliability and agreement among people perceptions of event boundaries. Mere task experience appears sufficient to identify basic events in a novel domain, although the same is not true of determining proper display choice.

The above results should not be taken to indicate that subjects are slaves to the model they trained under and the task environment. While subjects did converge to the display model they were trained under, subjects in the expert condition appeared to generalize their knowledge broadly. These subjects showed increased convergence over time to the second stage of all three expert models. In other words, exposure to one expert's view of the world encouraged more general expert-like behavior, rather than behavior that was only closely coupled to the particular training model. Control subjects did not show this systematic improvement in fit to all expert models. While control subjects might display idiosyncratic behaviors that agree with one expert model, they did not learn behaviors that were consistent across experts.

Overall, our results suggest that related training methods should prove successful in expediting the transition from novice to expert-levels of performance. Using an expert's RADAR model to train novices sidesteps several thorny issues. RADAR's fit of an expert quantifies the expert's action patterns (avoiding the limitations and effort involved in self-report) and provides a means to communicate this expertise to a novice in a task-situated manner.

There is a lot more research to be done before such training methods can be perfected. Although not reported above, expert RADAR models differed greatly in how well they fit each subject. One expert model fit particularly well, achieving the best fit for 18 of the 30 subjects. Interestingly, the model of the highest performing expert fit the subjects in our study the worst, with only 5 subject being well fit by it. One important challenge is determining which expert model is most beneficial for each novice at each stage in training. One possibility is that novices will vary in terms of which expert model is best. Hopefully, RADAR's formal approach will allow for best practices to be determined.

Acknowledgments. This research was supported in part by AFOSR Grant FA9550-07-1-0178 and NSF CAREER Grant #0349101 to B.C.L.

References

1. McTear, M.F.: User modeling for adaptive computer systems: a survey of recent developments. Artifcial Intelligence Review 7, 157–184 (1993)
2. Goodman, B.A., Litman, D.J.: On the interaction between plan recognition and in-telligent interfaces. UMUAI 2, 83–115 (1992)
3. Yi, W., Ballard, D.: Behavior recognition in human object interactions with a task model. In: AVSS, pp. 64–64 (2006)
4. St. John, M., Smallman, H.S., Manes, D.I.: Assisted focus: Heuristic automation for guiding users' attention toward critical information. In: Proc. of the 1st Inter. Conf. on Augmented Cognition (2005)
5. Horvitz, E., Barry, M.: Display of information for time-critical decision making. In: Proc. of Conf. on Uncertainty in AI, pp. 296–305 (1995)
6. Mantyjarvi, J., Seppanen, T.: Adapting applications in mobile terminals using fuzzy context information. In: Mobile HCI, pp. 95–107 (2002)
7. Schneider-Hufschmidt, M., Kuhme, T., Malinowski, U.: Adaptive User Interfaces: Principles and Practice. North-Holland, Amsterdam (1993)
8. Albrecht, D.W., Zukerman, I., Nicholson, A.E.: Bayesian models for keyhole plan recognition in an adventure game. User Modeling and User-Adapted Interaction 8(1-2), 5–47 (1998)
9. Teevan, J., Dumais, S.T., Horvitz, E.: Personalizing search via automated analysis of interests and activities. In: Proceedings of SIGIR, pp. 449–456 (2005)
10. Jones, G.J.F., Brown, P.J.: Context-Aware Retrieval for Ubiquitous Computing Environments. Springer, Heidelberg (2004)
11. Gajos, K.Z., Everitt, K., Tan, D.S., Czerwinski, M., Weld, D.S.: Predictability and accuracy in adaptive user interfaces. In: Proceeding of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1271–1274 (2008)
12. Love, B.C., Jones, M., Tomlinson, M.T., Howe, M.: Learning to predict informa-tion needs: Context-aware learning to predict information needs: Contex-aware display as a cognitive aid and an assessment tool. In: Proceedings of the 27th annual SIGCHI conference (2009)
13. Biederman, I., Shifrar, M.: Sexing day-old chicks: A case study and expert sys-tems analysis of a difficult perceptual learning task. Journal of Experimental Psychology: Learning, Memory, & Cognition 13, 640–645 (1987)
14. Feldon, D.F.: The implications of research on expertise for curriculum and pedagogy. Educational Psychology Review 19, 91–110 (2007)
15. Hoffman, R.R., Crandall, B., Shadbolt, N.: Use of the critical decision method to elicit expert knowledge: A case study in the methodology of cognitive task analysis. Human Factors 40, 254–276 (1998)
16. Ark, T.K., Brooks, L.R., Eva, K.W.: The benefits of flexibility: The pedagogical value of instructions to adopt multifaceted diagnostic reasoning strategies. Medical Education 41, 281–287 (2007)
17. Kluger, A., DeNisi, A.: Effects of feedback intervention on performance: An his-torical review, a meta-analysis, and a preliminary feedback intervention theory. Psychological Bulletin 119, 254–284 (1996)

18. Gonzalez: Decision support for real-time, dynamic decision-making tasks. Organizational Behavior and Human Decision Processes 96, 142–154 (2005)
19. Ashby, F., Maddox, W.: Human category learning. Annual Review of Psychology 56, 149–178 (2005)
20. Saffran, J.R.: Statistical language learning: Mechanisms and constraints. Current Directions in Psychological Science 12, 110–114 (2003)
21. McRae, K., Spivey-Knowlton, M.J., Tanenhaus, M.K.: Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. Journal of Memory and Language 38, 283–312 (1998)
22. Kurby, C.A., Zacks, J.M.: Segmentation in the perception and memory of events. Trends in Cognitive Sciences 12, 72–79 (2008)
23. Zacks, J., Speer, N., Vettel, J., Jacoby, L.: Event understanding and memory in healthy aging and dementia of the Alzheimer type. Psychology & Aging 21, 466–482 (2008)
24. Speer, N., Swallow, K., Zacks, K.: Activation of human motion processing areas during event perception. Cognitive, A_ective & Behavioral Neuroscience 3, 335–345 (2003)
25. Newton, D.: Foundations of attribution: The perception of ongoing behavior. In: Harvey, J. (ed.) New Directions in Attribution Research, Erlbaum, Mahwah (1976)
26. Cleeremans, A.: Mechanisms of implicit learning: Connectionist models of sequence processing. MIT Press, Cambridge (1993)
27. Gureckis, T., Love, B.C.: Transformational vs. statistical complexity in sequential learning. Cognitive Science (in press)

Translating Learning Theories into Physiological Hypotheses

Jennifer J. Vogel-Walcutt[1], Denise Nicholson[1], and Clint Bowers[2]

[1] Univeristy of Central Florida, Institute for Simulation and Training, 3100 Technology Parkway, Orlando, Florida, 32826
[2] Univeristy of Central Florida, Department of Psychology, 4000 Central Florida Boulevard, Orlando, Florida, 32816
jjvogelwalcutt@yahoo.com, dnichols@ist.ucf.edu, bowers@mail.ucf.edu

Abstract. The battlefield has become an increasingly more complicated setting in which to operate. Additional stressors, complexity, and novel situations have challenged not only those in the field, but consequently also those in training. More information must be imparted to the trainees, yet more time is not available. Thus, in this paper, we consider one way to optimize the delivery and acquisition of knowledge that can be meaningfully applied to the field setting. We hypothesize that for learning efficiency to be maximized, we need to keep learners in a constant state of engagement and absorption. As such, we consider neuro-physiological hypotheses that can help prescribe mitigation strategies to reduce the impact of sub-optimal learning.

Keywords: Learning efficiency, Augmented cognition, Adaptive training.

1 Introduction

The battlefield has become an increasingly more complicated setting in which to operate. Additional stressors, complexity, and novel situations have challenged not only those in the field, but consequently also those in training. More information must be imparted to the trainees, yet more time is not available. To accomplish this goal, researchers not only need to consider the type of information imparted to the trainees, but also the speed of delivery, the rate of acquisition, and the time required to reach fully-trained status. Thus, in this paper, we consider one way to optimize the delivery and acquisition of knowledge that can be meaningfully applied to the field setting. We hypothesize that for learning efficiency to be maximized, we need to keep learners in a constant state of engagement and absorption. In response to this goal, current research has identified many issues that can affect learning, including but not limited to boredom, arousability, anxiety, stress, working memory overload, and workload.

To deal with these issues, researchers have investigated mitigation strategies to lessen their affect, including principles from cognitive load theory (CLT) [25], information presentation strategies [12], feedback, scaffolding [17], working memory supports, and others. Today, these strategies are applied based on general principles

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 678–686, 2009.

[6] [12]across groups of people. Our goal, however, is to individualize the application of strategies in an effort to increase effectiveness by taking into account individual differences. This way, we can apply them only when an individual needs them and we can apply only those specific strategies needed by an individual [29]. To accomplish this, we are investigating the use of neuro-physiological data to define when a learner is in a sub-optimal learning state.

Thus our ultimate goal is to identify these sub-optimal states and prescribe strategies to mitigate their negative impact on learning. To do this, we must first identify those states that most affect learning and then generate hypotheses about how these states may be optimized by changes in the learning environment. In this paper, we investigate three states that are well supported in the literature to affect learning and begin to identify possible strategies to decrease their impact. Our three main areas of consideration include workload, arousal, and boredom.

2 Workload

First, we consider workload. Workload is a multidimensional construct that describes the level of effort expended to complete a task. Some of the dimensions include temporal, spatial, physical, and mental workload [9]. When we apply the concept of workload to learning tasks, we typically refer to mental or cognitive workload. To determine its impact on learning, we consider Cognitive Load Theory (CLT) [25]. CLT posits that there is a limited working memory capacity and when that capacity is exceeded, information is lost. In other words, when workload levels are too high or too low, learning performance is hindered. Thus, during supra-optimal cognitive load conditions, the learner must choose the information to which they will attend and the remaining information is ignored. This is particularly problematic in the case of novice learners because the novice learner is unable to determine the difference between relevant and irrelevant information. Consequently, they will commonly shed information that is important to achieve the learning goals.

Low workload can also be problematic as it leads to wayward attention. When the learner is operating under limited or sub-optimal workload levels they begin to attend to other information aside from the intended learning material. This is particularly problematic if the extraneous information becomes the primary focus of learner attention. As a result, germane information to the learning experience is lost, and learning suffers. However, when workload levels are optimal, a learner performs best. The combination of mental, physical, and temporal demands is in balance, allowing an individual to maintain focus on a task while not being overloaded.

CLT suggests that workload can be further divided into three different types, namely germane, extraneous, and intrinsic [13]. Intrinsic cognitive load is that load which is inherent in the learning material, whereas germane cognitive load is the effort the learner expends to acquire, assimilate, and file that information into long-term memory. Finally, extraneous cognitive load is the load that is imposed upon the learner and detracts from the learning experience. Currently, to measure workload we use either self-report measures (TLX [9]; MRQ [4]; & CLQ [14]), or neuro-physiological measures [3], inter-heart rate variability, and pupil dilation [24], and to date, we cannot distinguish between the three types of cognitive load. One additional problem with all these measures is that they can only tell us what to do in a subsequent task. In other words, if workload is too high we can subsequently *do less* and when it is too low we simply *do more*. However, what is more likely the case is that workload varies throughout a learning experience. In other words, there are times during the learning experience when the learner is operating with too much cognitive load and other times when there is a sub-optimal level of cognitive load. Their levels are inconstant flux. Assuming this is the case, then there are times during the learning episode when the learner could benefit from additional instructional prompts or secondary tasks but, at the same time, there are situations under which those same interventions would be problematic.

Today, because we can only consider the average workload levels across the whole learning cycle, interventions intended to mediate workload are not nearly as precise as those that could be provided with real-time measurements. These average scores allow us to make few, and likely not very impactful, changes to subsequent learning experiences. Thus, we have two goals to improve augmentation of workload. First, we need to identify workload levels in real-time and second, we need to be able to prescribe changes to the learning environment immediately. Because efforts to achieve the first goal are already underway [27], we focus our attention on the second goal which is to determine what to do with the information once we have it. In this paper, we aim to determine what we might do to mitigate sub-optimal workload conditions to improve learning. To that end, we offer two hypotheses.

2.1 Decrease Workload

We hypothesize that when workload is high, the learner is overwhelmed by the incoming information and mitigation strategies should be applied to reduce task-shedding. To mitigate these effects, we consider strategies that help reduce the complexity of the scenario, provide "cheat sheets" or guides to reduce the strain on working memory, and identify the most germane elements of the training material in order to focus the learner's attention. More specifically we hypothesize that when workload is too high during a learning experience, information should be presented to the learner in such a way that it first focuses their attention on the relevant information and second, reduces the drain on working memory capacity. For example, if we applied this hypothesis to simulation-based training, the most commonly used platform for training and assessing applied, field-based knowledge in the military, we would state:

When a learner's workload is too high, providing full-screen instructional prompts will allow the learner to focus exclusively on the prompt and to avoid distraction by the overwhelming learning stimuli. Consequently, the learner's focus on the relevant information needed to achieve the learning goal is increased, positively affecting learning efficiency.

2.2 Increase Workload

Conversely, when workload is too low the learner is at risk of being distracted by extraneous, irrelevant information. In this case, the use of an instructional prompt is also warranted, but the goals are different than when workload is too high. Instead of reducing the information on the screen in order to focus the learner's attention, we increase the information on the screen and allow them to learn multiple pieces of information simultaneously. Providing a secondary task to learners whose workload levels are too low may increase their workload levels to an optimal state. This acts as a secondary task. Increasing workload levels may further increase learning efficiency because the learner is allowed to more quickly, or more efficiently, progress through the learning material. If we again apply this hypothesis to simulation-based training we would hypothesize:

When a learner's workload levels are too low, provide a scaffolded instructional prompt in addition to the present learning material. As a result, workload levels will increase and learning efficiency will improve.

3 Arousal

Arousal is another construct for consideration. While there remain many different conceptualizations in the literature, for our purposes we use the definition from Schatz (2009) where general arousal refers to a person’s physical activity, alertness, and emotionality. Theoretically, there is an optimal state of arousal where learners can best acquire, encode, and store information [19]. In this case, a mid level of arousal leads to better attention. So while arousing stimuli can attract attention, it can also distract the learner. As such, similar to workload, when arousal levels are too high or too low, learning decrements can occur. The mechanism by which this works is that when learners are experiencing a high or low arousal state, they must use some of their working memory capacity to either deal with the overwhelming stimuli or to maintain alertness. As such, attention is strained and eventually wanes [10] [19]. When a person is under-aroused, their processing ability diminishes and their attention narrows [19] whereas when their arousal is too high, their attention focuses on declarative knowledge, reducing focus on comprehension of the material and subsequently impeding the learner’s ability to apply that knowledge to real-world settings [7]. Thus, with arousal, we consider both the mitigation of under as well as over arousal.

3.1 Decrease Arousal

Specifically, during high arousal states learners will select, often haphazardly, particular information on which to focus, ignoring other stimuli and ultimately resulting in task-shedding [26]. In this case, the learner fails to acquire the necessary learning material. Typically, when this happens the learner chooses to attend to the easiest, or most explicit, information and consequently ignores more meaningful or implicit information. As a result, high arousal leads to low initial recall but higher long-term retention for procedural skills and declarative knowledge [23] yet it inhibits long-term retention of semantic or meaningful information [20] [7]. If it were the case that we were interested solely in procedural skills or declarative knowledge, we might prefer to keep learners in a high arousal state. However, for information to be useful for the military, it must be meaningfully acquired so that it can be applied to the field setting. To decrease arousal we consider reducing “noise” (extraneous load), providing working memory aids [29], and scaffolding the learner's training at a slower pace [17]. Specifically, when we apply these concepts to simulation-based training we hypothesize the following:

When arousal levels are too high, reducing the complexity of the given scenario while simultaneously providing directive instructional prompts, will help the learner better focus their attention. Reducing the number of overwhelming stimuli and scaffolding the learner's knowledge to acquire meaningful knowledge will lead to more effective and efficient training.

3.2 Increase Arousal

During low arousal states, reduced learning can occur because the learner must work hard to maintain attention. Fatigue and the reduced availability of working memory capacity can interfere with learning. Thus, while low arousal states can lead to better short-term memory, long-term memory may suffer [19]. As such, a mid level of arousal is preferred. Mitigation strategies for increasing arousal include adding secondary tasks, increasing complexity, progressing the learner, and learning two types of material simultaneously. Again if we apply these concepts specifically to simulation based training, we offer the following:

When a learner is under aroused during a scenario, two changes should be made. First, increase the complexity of the scenario so as to capture the attention of the learner and second, provide instructional prompts to the learner at a faster rate so as to more quickly scaffold their learning. As a result, the likelihood that the learner will focus attention on the learning material and more quickly acquire knowledge is increased. Consequently, learning effectiveness and efficiency will improve.

4 Boredom

Expanding the concept of arousal we consider the effect of arousal combined with affect and how they may lead to feelings of boredom. Boredom is a construct that has been widely discussed across disciplines, yet has not been so widely studied, either systematically or empirically. Thus, while the limited research in this area supports the hypothesis that boredom highly and negatively impacts learning, accurate, empirically tested, solutions to mitigate the impact of boredom, have not yet been well investigated [2]. Several different definitions of boredom exist in literature, yet a general consensus has yet to be achieved [5] [8] [15] [16] [18] [31]. For the purposes of our paper, we define boredom as the combination of low engagement with negative affect. We considerate it for this paper because it is often professed to be one of the major issues leading to reduced learning, wayward attention, and a lack of motivation [1]. Learners experiencing boredom must expend conscious efforts to maintain attention and eventually their efforts fail resulting in low engagement and, consequently, learning suffers [8]. Further, unmotivated students spend approximately only 42% of their time on task, which also leads to a reduction in acquisition of knowledge [22]. Thus, to combat the effects of boredom we must address both areas of concern. First, we need to improve engagement and second, we need to improve affect.

We hypothesize that when engagement and workload levels are low and performance is low, the student is bored and mitigation strategies should be applied to capture the learner's attention. To mitigate these issues, we consider several strategies, generally focused on information presentation. Specifically, Mayer's principles (2005) for multimedia may help address some of the issues caused by feelings of boredom. Further, interest is the opposing force to boredom and as it increases, boredom decreases. Topics such as relevancy, presentation type, and challenge have been suggested to influence interest [11] [21].

4.1 Meyer's Principles

First, we consider Mayer's (2005) multimedia principles to help guide when and how to present information to learners in the most effective way possible. Based on CLT, Mayer applied their theoretical principles to multimedia presentations in an effort to improve learning efficiency. While he has defined and studied numerous principles, we consider six that are most relevant to increasing engagement levels, one aspect of boredom. Specifically these include the principles of multimedia, split-attention, redundancy, personalization, guided-discovery, and self-explanation. The principle of multimedia suggests that providing information in two different formats can improve learning efficiency. Split-attention occurs when the learner must attend to two competing modes of presentation such as narration and text simultaneously. The learner is forced to split their attention because both tax the auditory channel. The principle of redundancy refers to information being presented in excess of that which is necessary to achieve the learning goals. Because this information detracts the learner from the relevant learning material, it reduces learning efficiency. The principle of personalization requires that the learning material be personally relevant to the learner thus increasing their interest in, and subsequently their motivation to learn, the material at hand. Guided-discovery involves scaffolding the learner through a learning experience that helps them distinguish relevant from irrelevant information. Finally, self-explanation occurs when the learner is expected to explain or describe the learning material. It is the difference between having to acquire knowledge and having to apply knowledge.

So how do these principles help mitigate the effects of boredom on learning? Utilizing the principles set forth by Mayer may allow us to improve the level of the learner's engagement and by doing so, we address one prong of the boredom construct. Thus, based on Mayer's work, we recommend the following:

1. Multimedia presentations should use colloquial language. This makes it easier for it the learner to follow the information being presented.
2. The learning material should be narrated. This allows the learner to be able to both attend to pictorial presentations while simultaneously using the auditory channel to take in verbal information without having to split attention.
3. Information should be provided in two modes. Typically this involves pictures and narration.
4. Pictures and narration should be provided simultaneously. This allows for the most efficient delivery of information.

5. No extraneous information should be included. Again, this helps the learner, especially the novice learner, focus their attention on only that information that is relevant to the learning goals.

Thus, if we apply these principles to simulation based training to make more specific hypotheses, we provide the following:

When engagement in the learning material is low, using multimedia presentations involving both pictures and narration, using colloquial language, and reducing the amount of extraneous information presented will lead to more effective and efficient learning. By capturing and maintaining the learner's attention, their engagement in the learning material will increase leading to improved learning effectiveness and efficiency.

4.2 Interest

The second prong of boredom is negative affect. Here we consider the concept of interest as it is the opposite of boredom. High interest leads to high engagement which, in turn, leads to high motivation. When motivation is high, we expect deeper learning and we also expect learning efficiency to increase. To improve interest, we consider the work of Vogel-Walcutt, et al. (2009) who identified seven categories of interventions. Specifically, they focused on relevancy, meaning, control, collaboration, change, involvement, and creativity. Relevancy refers to relating the learning material to personal interests. Control allows learners to take ownership in the learning experience, while collaboration requires them to work with others in a team. Change or novelty typically involves either surprises or a change of pace in the presentation of the learning material. Involvement refers to keeping the learner at an optimal level of stimulation utilizing strategies such as competition. Creativity typically requires the learner to work with pictures or in a story format to utilize other skills and focus on meaningful knowledge. Finally, to create meaning, it is necessary to focus on applied learning or using a problem-solving approach. Before we can apply these strategies in real-time, we must operationalize interest based on neurophysiological measures. To date, engagement levels, which can be measured by an electroencephalogram (EEG), provide our best information. However, additional work is being conducted to address the need to also quantify affect in real-time [28]. Again, applying these principles to simulation based training, we consider the following hypotheses:

When interest is low so, too, is engagement and consequently motivation. Thus, when interest in the material is low, as defined by engagement, allowing learners to change the material so that they may be able to make it more personally relevant and also allowing them to compete with others may increase interest and drive motivation to learn. Ultimately, this results in more effective and efficient learning experiences.

5 Summary

Thus, in this paper, we consider neuro-physiological hypotheses that can help prescribe mitigation strategies to reduce the impact of sub-optimal learning. We provide recommended strategies, based on current research, to help achieve the goal of keeping learners in a constant state of engagement and absorption. Our overarching goal is to individualize the application of mitigation strategies, based on individual differences, in order to increase learning efficiency in complex real-world military training settings.

Acknowledgement

This work is supported in part by the Office of Naval Research Grant N000140710098. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the ONR or the US Government. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

1. Anderson: Student engagement and dropout: An investigation with students who have mild disabilities. Dissertation, University of Minnesota (2004)
2. Belton, T., Priyadharshini, E.: Boredom and schooling: A cross-disciplinary exploration. Cambridge J. of Ed. 37(4), 579–595 (2007)
3. Berka, C., Levendowski, D., Cvetinovic, M., Petrovic, M., Davis, G., Lumicao, M., Zivkovic, V., Popovic, M., Olmstead, R.: Real-time Analysis of EEG Indices of Alertness, Cognition, and Memory with a Wireless EEG Headset. International Journal of Human-Computer Interaction 17(2), 151–170 (2004)
4. Boles, D.B., Bursk, J.H., Phillips, J.B., Perdelwitz, J.R.: Predicting dual-task performance with the Multiple Resources Questionnaire (MRQ). Human Factors 49(1), 32–45 (2007)
5. Chen, A.: Perception of boredom: Students' resistance to a secondary physical education curriculum. Research in Middle Level Education Quarterly 21(2), 1–20 (1998)
6. Clark, R.C., Nguyen, F., Sweller, J.: Efficiency in learning: Evidence-based guidelines to manage cognitive load. Pfeiffer, San Francisco (2006)
7. Eysenck, M.W.: Arousal, learning, and memory. Psych. Bul. 83(3), 389–404 (1976)
8. Fisher, C.: Boredom at work: A neglected concept. Human Relations 46(3), 395–417 (1993)
9. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, pp. 239–250. North Holland Press, Amsterdam (1988)
10. Kahneman, D.: Attention and effort. Prentice-Hall, Englewood Cliffs (1973)
11. Kopp, T.: Designing boredom out of instruction. Perf. & Instr. 21(4), 23–27 (1982)
12. Mayer, R.E. (ed.): The Cambridge handbook of multimedia learning. Cambridge University Press, New York (2005)
13. Paas, F., Tuovinen, J.E., Tabbers, H., Van Gerven, P.W.M.: Cognitive load measurement as a means to advance cognitive load theory. Ed. Psyc. 38(1), 63–71 (2003)
14. Paas, F.G.W.C.: Training strategies for attaining transfer of problem-solving skill in Statistics: A cognitive-load approach. Journal of Educational Psychology 84(4), 429–434 (1992)
15. Pekrun, R., Elliot, A.J., Maier, M.A.: Achievement goals and discrete achievement emotions: A theoretical model and prospective test. J. of Ed. Psych. 98(3), 583–597 (2006)
16. Perkins, R.E.: The nature and origins of boredom. Dissertation (1981)

17. Puntambekar, S., Hübscher, R.: Tools for scaffolding students in a complex environment: What have we gained and what have we missed? Educational Psychologist 40(1), 1–12 (2005)
18. Rude, R.: Isn't that interesting! Ed. Week 20(43), 47–49 (2001)
19. Schatz, S., Bowers, C.A., Nicholson, D.: Arousability and learning: Arousability may moderate workload and anxiety during high-simulation instruction. To be submitted to Learning and Individual Differences (2009) (Manuscript in preparation)
20. Schwartz: Individual differences in cognition: Some relationships between personality and memory. J. of Res. In Personality 9(3), 217 (1975)
21. Schweinle, A., Meyer, D.K., Turner, J.C.: Striking the right balance: Students' motivation and affect in elementary mathematics. The J. of Ed. Res. 99(5), 271–293 (2006)
22. Shernoff, D.J., Csikszentmihalyi, M., Schneider, B., Shernoff, E.S.: Student engagement in high school classrooms form the perspective of flow theory. School Psych. Quart. 18(2), 158–176 (2003)
23. Steidl, S., Mohi-Uddin, S., Anderson, A.K.: Effects of emotional arousal on multiple memory systems: Evidence from declarative and procedural learning. Learning and Mem. 13, 650–658 (2006)
24. St. John, M., Kobus, D., Morrison, J.G., Schmorrow, D.: Overview of the DARPA augmented cognition technical integration experiment. Int. J. of Human-Computer Interaction 17(2), 131–149 (2004)
25. Sweller, J.: Cognitive load during problem solving: Effects on learning. Cognitive Science 12, 257–285 (1988)
26. Tsang, P., Wilson, G.F.: Mental workload. In: Salvendy, G. (ed.) Handbook of Human Factors and Ergonomics, pp. 417–489. John Wiley & Sons, Inc, New York
27. Vartak, A.A., Fidopiastis, C.M., Nicholson, D.M., Wafsy, M.B., Schmorrow, D.D.: Cognitive state estimation for adaptive leaning systems using wearable physiological sensors. In: The proceedings of Biosignals 2008, Funchal-Madeira, Portugal, January 28-31 (2008)
28. Vartak, A., Fidopiastis, C., Nicholson, D., Mikhael, W.: Estimation of arousal using decomposed skin conductance features. In: Biomedical Sciences Instrumentation (2009) (accepted for publication)
29. Vogel-Walcutt, J.J., Schatz, S., Bowers, C., Gebrim, J.B., Sciarini, L.W., Nicholson, D.: Augmented cognition and training in the laboratory: DVTE system validation. Human Factors and Ergonomics Society Annual Meeting Proceedings 52(3), 187–191(5) (2008)
30. Vogel-Walcutt, J.J., Bowers, C.A., Marino-Carper, T., Nicholson, D.: Increasing Learning efficiency in military learning: More than just working memory (2009) (manuscript in preparation)
31. Watt, D.: Fighting more than fires: Boredom-proneness, workload stress, and underemployment among urban firefighters. DAI: Section B: The Sciences and Engineering US: Univ Microfilms International 63(5-B), 2637 (2002)

Adapting Instruction

Wallace H. Wulfeck II

Space and Naval Warfare Systems Center Pacific
San Diego, California
wally.wulfeck@navy.mil

Abstract. It is often claimed that adapting instruction to an individual's progress, personal characteristics or preferences will somehow increase learning, and there have been literally thousands of studies over at least the past 100 years exploring this idea. This presentation will review various approaches to adapting instruction, such as changing the rate, difficulty, sequence or structure, instructional strategy or instructional media on the basis of learner progress, prior knowledge, aptitudes or preferences, under various forms of instructor, learner, program, or opponent control. This paper gives an organizing framework, describes some of the theoretical underpinnings for particular adaptations, and describes experimental and practical criteria for evaluating claims of efficacy and efficiency of instructional adaptations.

1 Introduction

We have known for over 2000 years how to provide effective instruction to an individual: provide expert tutors who personally coach the learner in a shared workspace over a period of years. The tutors continuously adapt instruction to the individual, structure the subject matter to be learned, decide on activities to engage and motivate the student, determine the sequence of topics to be learned or tasks to be mastered, provide instruction and context, model correct performance through worked examples, focus practice, observe the student's performance, monitor progress, correct or remediate performance lapses, provide feedback and coaching for improvement, conduct dialogues about the task and provide amplification and meta-instruction. This was the model of instruction at the time of Aristotle and Alexander and it is still used in apprenticeships, coaching, and in graduate-level and professional education.

The problem, of course, is that one-on-one tutoring is expensive. In the last few hundred years, instruction for groups of students became more common, in order to achieve efficiency in mass education. In the 20th century, perhaps as a result of the industrial revolution, a view of instruction as assembly-line manufacturing emerged, and attempts were begun to automate instructional processes. Standardized testing began during the First World War, "teaching machines" were invented in the 1920s, instructional films in the 1940s, programmed instruction in the 1950s, and with the development of television and early computers came "educational technology" in the 1960s and 1970s.

Although the usual goal of these instructional technologies is merely to make the delivery of instruction less expensive but not less effective than conventional

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 687–695, 2009.

instruction, sometimes a stated goal is to achieve learning gains in mass instruction comparable to those routinely achieved by one-on-one tutoring. Benjamin Bloom [3] famously described the "two-sigma problem," noting that group instructional methods rarely achieve gains in performance comparable to the two-standard-deviation gains that are often observed in tutoring. In general, the idea is that if instruction could somehow be more responsive to individual students, that is adapted to individuals rather than to (the slowest members of) a group, then learning might be more efficient or effective.

2 Adapting Instruction

A modern general conceptualization of the control of instruction was first mapped out by Smallwood [29] in the context of control theory, and substantially formalized by Atkinson [1] in a now-classic paper describing "ingredients for a theory of instruction." In Atkinson's view, control of instruction, and therefore of adaptation involves (1) a model of the learning process, (2) specification of admissible instructional actions, (3) specification of instructional objectives, and (4) a measurement scale for costs of actions and payoffs of achievement of objectives.[1] The model of the learning process requires a reasonably precise characterization of an individual's state of learning before and after an instructional action, and therefore, by implication, some measure of the student. More generally, characterizing the model of learning and the effect of instructional actions on learning in any precise way really requires some sort of theory of learning and/or behavior change.

There are many ways to modify or adapt instruction, and they may be mixed. Following Atkinson's [1] characterization, we will describe some dimensions of adaptation in terms of models of learning, instructional actions, and implementation of control. In some cases there may be similarities or overlap across these dimensions.

2.1 Basis for Adaptation

Perhaps the most important dimension of adaptation is the characterization or model of the learning process. The model specifies some properties of the student's learning process which we will observe and use as a basis for prescribing some instructional action or treatment. This at least requires some measurement of the individual, together with the theoretical expectation that variation in that measurement has some instructional importance. That is, observed differences for individuals on that measure will lead to differences in whether and how we will instruct, and therefore to a difference in outcome. Among the properties of individuals and/or their learning or performance to which adaptation may occur are at least the following:

---

[1] Here the term "objective" refers to a goal or overall outcome of instruction, not a "behavioral objective" in the sense of Gagne or Mager. Atkinson (1972) gave examples of alternative instructional goals such as: (1) maximize the mean performance of the whole class, (2) minimize the variance in performance for the whole class, (3) maximize the number of students who score at grade level, or (4) maximize the mean performance for each individual.

Individual preference. The easiest way to get a measure from individuals is to simply ask them whether they feel that they know a concept, or are ready to take a test, or are ready to move on, or need additional explanation or practice, or prefer one or another style of instructional presentation (Also, see *learner control* below.) Unfortunately, it has been known for at least half a century that students are not particularly good judges of their own states of knowledge [1] and therefore not very good at judging their needs for explanation, practice, or remediation.

Current progress / result / score on some measure. This may be as simple as whether the last answer to a probe test question was correct or incorrect, or some estimate of the likelihood that a student has learned a particular concept, or it may be a more complex measure, such as a learning rate or a pattern of responses. A very common adaptation strategy had its roots in early programmed instruction (cf. Skinner, 1954): present a "frame" or block of instruction which concludes with a question. If the response is correct, proceed to the next frame; otherwise inform the student of an error and present the frame again. This style is still used today in powerpoint-derived web-based training because it is easy to implement, if not very effective. Much more sophisticated approaches have been based on mathematical models of learning (cf. 1), or extensive cognitive analyses of tasks [e.g. 4; 2].

Traits or aptitudes. It is often claimed that different individuals have different learning styles or aptitudes, and that adapting instruction to them will somehow increase learning. There have been literally thousands of studies over at least the past 100 years exploring this idea. Proponents of one or another instructional method generally seize upon one supposed characteristic (usually a binary one) that might differentiate individuals, and then propose different instructional treatments depending on the individual's classification according to that characteristic. For example, we see learners described as:

- Right-brain vs. Left-brain
- Active vs. Passive
- Wholist vs. Serialist
- Visual vs. Auditory
- Multi-tasking vs. Sequential
- Abstract vs. Concrete
- Convergent vs. Divergent
- Extravert vs Intravert
- Type A vs Type B
- Sensing vs. Intuitive
- Thinking vs. Feeling

A quick tour through Wikipedia starting with topics such as "Learning Styles" or "Individual Differences" will lead to dozens of these sorts of dichotomies. More sophisticated theorists will combine two or more of these dichotomies into multi-dimensional constructs. Unfortunately, these oversimplified approaches rarely work.

An adaptation to an individual difference actually requires substantial analysis and experimentation in order to demonstrate effectiveness. First, there must be some observable difference between individuals. Preferably there should be some reasonably scientific theory about why the difference exists and why adaptation to the difference should have any effect on an instructional outcome. Second, there must be some sort of test or measurement for the difference that reliably differentiates individuals. The key here is “reliably”. Further it must be possible to measure this difference quickly enough to affect instructional decision making. Also, the difference must be stable long enough for an instructional strategy to be applied and take effect. Then there must be some prescription for an instructional adaptation to the difference, and there must be a reliable post-instruction measurement so that differential outcomes could be observed. Finally, there must be an experimental or quasi-experimental evaluation of the effectiveness of the adaptation. Further, results of that evaluation ought to show a disordinal interaction between the individual difference variable and the adaptation variable.

Despite hundreds of studies over 40 years of research on what are now called “aptitude-treatment interactions”, and several major reviews [13; 12; 33, 30, 37] there is almost no evidence of effective and practical adaptation of instruction to individual traits or aptitudes.

According to Tobias [35]: “Instructional designers are often urged to adapt instruction to students’ learning styles. The persistence of the learning style concept is amazing—a testament to the gullibility of even well-informed individuals who ought to know better. It seems that advocates of learning styles have never heard of the history of ATI research, which attempted to provide a database for adapting instruction to student characteristics and found many thorny problems. It is probably fair to say that the popularity of adapting instruction to learning styles is matched only by the utter absence of support for this idea.”

Prior knowledge / skill or ability level / prior achievement. In contrast to the dismal results for aptitude-treatment interactions, good results have been observed when adapting instruction to prior achievement. It has been shown that students with low ability or low knowledge of a particular topic generally need increased instructional guidance or support for learning [33, 34, 36, 23]. Also, more complex or broader measures are often used for more macro-level instructional decisions. For example, completion of pre-requisites is often a condition for admission to an advanced course, or we may decide whether a student is ready for a particular topic based on an aggregate measure of prior achievement or prior knowledge.

2.2 Instructional Actions for Adaptation

Rate or Pacing. Instruction may slow down or speed up, for example by allowing more or less time for study on a topic, or by providing more or fewer examples or practice opportunities. Fletcher [16] notes that individualization of pace is by far the most common adaptation to individual differences in learning, and is so common that it is “frequently, although incorrectly, treated as synonymous with individualization of instruction.” Fletcher [18] describes a “Rule of Thirds” which is a statistical summary of meta-analyses of the use of computer-based adaptation of pacing: it “can either reduce instructional time to reach instructional goals by about one-third (a goal more characteristic of training than education), or increase the skills and knowledge acquired by about one-third while holding instructional time constant (a goal more characteristic of education than training)."

Difficulty. Chunks of instructional content may be graded or calibrated for difficulty or ease of learning, and more or less difficult concepts may be presented. Topics may be skipped or reinforced. This approach is commonly taken in adaptive testing, where items are calibrated, then selected one by one, based on the correctness of previous answers, to converge quickly on an estimate of achievement or ability [cf. 38]. It is used less commonly to adapt instruction, because it is based on a normative approach while instructional content has structural relationships which can be exploited for sequencing.

Sequence or Structure. Instruction may be divided into separate topics on the basis of some logical order or content structure. Topics may be arranged in some heterarchy of superordinate-subordinate and relationship links, and instruction may proceed up, down, or sideways through the network. For example, topics thought to be prerequisite to others may be presented earlier, or conversely a holistic overview might precede more detail on particular topics. Surprisingly, however, changing the sequence of small chunks of instruction seems not to make much difference -- there were a number of studies in the 1960s that "scrambled" the order of frames in programmed instruction and found no differences in learning – unless there were strong dependency relationships among problems to be solved [39].

Method of Instruction. The instructional strategy may be varied. For example, a common instructional strategy called "Rule/Example/Practice" presents conceptual generalities or directions, followed by worked examples, followed by practice opportunity [cf. 24, 25]. Sometimes other orders of these components (e.g. "Example, Practice, Rule," 21) may even work better. An extension of this is a "problem-based" strategy that might start with an overall task or problem to be solved, and then work on separable parts of the problem while building knowledge.

Mode or Medium. The method of delivery of instruction may vary. For example, instructional materials might be presented in verbal or written form, or pictorially, or via a computer animation. In general however, results of research on differential effects of media are very mixed. Clark and Salomon [9] ask "why should we expect media to teach anyone anything?" and give references to many prior reviews of literature on media effectiveness. Clark [8] concludes "The best … evidence is that media are mere vehicles that deliver instruction but do not influence student achievement any more than the truck that delivers our groceries causes changes in our nutrition. Basically, the choice of vehicle might influence the cost or extent of distributing instruction, but only the content of the vehicle can influence achievement" (p. 13).

2.3 Control of Adaptation

Another dimension of adaptation involves the control of instructional alternatives, that is, how a control strategy for the instructional process is implemented.

Instructor control. Traditional instruction is usually directed and controlled by a teacher or tutor, who decides on rate of presentation, sequence, instructional strategy, and media, and most importantly, diagnosis, feedback, and remediation of performance problems and misconceptions.

Learner control. The learner might choose what topics to study, in what order, for how long, and may choose alternative delivery media or methods. There have been many studies of learner control, with very mixed results. Learners may be allowed to control the rate of instruction (often called self-pacing), the choice of instructional strategy or method, or control of sequence of topics or activities. In extreme forms, such as pure discovery or "constructivist" learning, students may explore instructional environments with little guidance. We referred above to Atkinson's [1] caution concerning self-pacing. In a broad review of learner control, Lunts [22] notes: "Thus, the research studies on LC fail to confirm or disconfirm anything. Consequently, there are no right answers on whether LC is beneficial for students and whether a higher degree of LC implied in a computer program improves instructional effectiveness" (pg 68). Finally, Mayer [23] concluded: "Pure discovery did not work in the 1960s, it did not work in the 1970s, and it did not work in the 1980s, so after these three strikes, there is little reason to believe that pure discovery will somehow work today" (pg 18). Kirshner, Sweller and Clark (2006) conclude that constructivist approaches to instruction are less effective than direct instruction.

Program or machine control. In computer-based instruction, programs can be written to choose instructional events, present and score practice or test items, provide written or pictorial content, and implement different instructional strategies. Investigations of the use of automation in instruction for the last 80 years have involved one form or another of program control. Its roots are in Pressey's [26] teaching machine, Skinner's (1954) programmed learning, and Crowder's [14] intrinsic programming. One approach, based on mathematical learning and memory theory and optimization was heavily investigated at Stanford University with good results [19]. Suppes, Fletcher & Zanotti [32] used extremely detailed analyses of mathematics curricula and research on mathematics learning to inform computer control of pace and sequence. Another approach is to design machine-based instructional systems based on analyses of tutoring. This approach began with early work by Carbonell, Collins, and colleagues. [cf. 7, 11, 31, 10] More recent approaches in intelligent tutoring systems involve extensive cognitive task analysis [cf. 28] and sophisticated logic for making control decisions [cf. 27]. The lesson learned from all of this work is that really deep content and cognitive analysis is necessary to construct effective instructional programs.

Opponent control. In competitive tasks, such as sports, war games, or business simulations, control of instructional events may be based on scripted scenarios, or at least on conventions. Ultimately, however, control of events may depend on what actions are chosen by an opponent. Often an instructor will act as an opponent in order to present instructive events or tune the level of difficulty or provide meaningful consequences to a trainee's actions. There is little research on the effectiveness of such strategies, although it is difficult to see how instruction and practice on competitive tasks could avoid opponent control completely.

2.4 Other Dimensions of Adaptation

In addition to the basis, actions, and control, there may be other variables that affect adaptation. One might be the type of content – it is possible that different models of learning might apply depending on whether one is learning facts such as in basic arithmetic versus high-level rules for problem solving. There have been many schemes for describing different types of content, but only some that then connect with alternative instructional actions or adaptations. Merrill's [24] approach is better than most.

Another dimension involves training individuals versus teams. Most tasks in the world of work actually involve teams or groups working together. Some development has been done on models of instruction for teams, or team instructional strategies. [e.g. 6], but very little has been done on adapting individual instruction within team training.

3 Conclusion

So, what we know at this point? Several points are worth making in summary:

First, individual tutoring is the most reliable way to achieve individualization and adaptation. When intelligent tutoring systems are carefully designed, they seem to achieve some of the same benefits. Good tutors adapt the pace, sequence of instruction, instructional strategy, amount of practice, and feedback. They do this by having a deep understanding or analysis of the content to be learned, strategies for explanation, and techniques for diagnosis, feedback, and remediation.

Second, computer-based adaptation of pacing based on careful models of learning and memory works and reliably yields about 30 percent time savings or increase in amount learned.

Third, both computer control of pacing/sequence and intelligent tutoring work best when there has been deep analysis and careful characterization of the instructional content and of the tutoring process. It is clearly time to revisit the learning trajectory optimization approaches developed at Stanford in the 1960s and 70s, and update them in light of more recent cognitive and content analysis techniques. More research and development on understanding and implementing in programs what tutors do to achieve Bloom's two-sigma difference in effectiveness is clearly warranted.

Finally, while this paper considered Atkinson's [1] first two ingredients for instruction, models and methods for adaptation, it has not discussed the third and fourth ingredients relating to instructional goals and the cost/effectiveness of actions to achieve them, mainly because so little solid work has been done. Fletcher [e.g. 15, 17] is one of the few voices in this wilderness.

References

1. Atkinson, R.C.: Ingredients for a theory of instruction. American Psychologist 27, 921–931 (1972)
2. Anderson, J.R., Corbett, A., Koedinger, K., Pelletier, R.: Cognitive Tutors: Lessons Learned. Journal of the Learning Sciences 4(2) (1996)
3. Bloom, B.S.: The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. Educational Researcher 13(6), 4–16 (1984)
4. Brown, J.S., Burton, R.R.: Diagnostic Models for Procedural Bugs in Basic Mathematical Skills. Cognitive Science 2(2) (1978)
5. Brown, J.S., Burton, R.R., de Kleer, J.: Pedagogical, Natural Language and Knowledge Engineering Techniques in Sophie I, II, and III. In: Sleeman, D., Brown, J.S. (eds.) Intelligent Tutoring Systems, Academic Press, London (1982)
6. Cannon-Bowers, J.A., Salas, E., Converse, S.A.: Cognitive psychology and team training: Training shared mental models and complex systems. Human Factors Society Bulletin 33, 1–4 (1990)
7. Carbonell, J.R., Collins, A.M.: Natural semantics in artificial intelligence. In: Proceedings of the IJCAI, pp. 344–351 (1973)
8. Clark, R.E.: Learning from Media. Information Age Publishers, Greenwich (2001)
9. Clark, R.E., Salomon, G.: Why should we expect media to teach anyone anything ? In: Clark, R.E. (ed.) Learning from Media, pp. 37–67. Information Age Publishers, Greenwich (2001)
10. Collins, A.M., Stevens, A.: A Cognitive Theory of Inquiry Teaching. In: Reigeluth, C.M. (ed.) Instructional design theories and models: An overview, pp. 247–278. Erlbaum, Hillsdale (1983)
11. Collins, A.M., Warnock, E.H., Passafume, J.J.: Analysis and synthesis of tutuorial dialogues. In: Bower, G.H., Spence, K.W. (eds.) The psychgology of learning and motivation, pp. 49–87. Academic Press, New York (1975)
12. Corno, L., Snow, R.E.: Adapting teaching to individual differences among learners. In: Wittrock, M.C. (ed.) Handbook of Research on Teaching, 3rd edn., pp. 605–629. Macmillan, New York (1986)
13. Cronbach, L.J., Snow, R.E.: Aptitudes and instructional methods: A handbook for research on interaction. Irvington, New York (1977)
14. Crowder, N.A.: Automatic tutoring by means of intrinsic programming. In: Galanter, E. (ed.) Automatic teaching: The state of the art, pp. 109–116. John Wiley, New York (1959)
15. Fletcher, J.D.: The effectiveness and cost of interactive videodisc instruction. Machine-Mediated Learning 3(4), 361–385 (1989)
16. Fletcher, J.D.: Individualized Systems of Instruction. In: Alkin, M.C. (ed.) Encyclopedia of Educational Research, pp. 613–620. Macmillan, New York (1992)
17. Fletcher, J.D.: Evidence for learning from technology-assisted instruction. In: O'Neil, Perez, R.S. (eds.) Technology applications in education, pp. 79–99. Lawrence Erlbaum Associates, Mahwah (2003)
18. Fletcher, J.D.: Remembrance of things past: The individualization imperative. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL (2007)
19. Fletcher, J.D., Atkinson, R.C.: An evaluation of the Stanford CAI program in initial reading (Grades K through 3). Journal of Educational Psychology 63, 597–602 (1973)

20. Gustafsson, J.-E., Undheim, J.O.: Individual differences in cognitive functions. In: Berliner, D.C., Calfee, R.C. (eds.) Handbook of Educational Psychology, pp. 186–242. Simon & Schuster Macmillan, New York (1996)
21. Lahey, G.F.: The effect of instructional sequence on performance in computer-based instruction. Journal of Computer-Based Instruction 7(4), 111–116 (1981)
22. Lunts, E.: What does the literature say about the effectiveness of learner control in computer-assisted instruction? Electronic Journal for the Integration of Technology in Education 1(2), 59–75 (2002), `http://ejite.isu.edu/Volume1No2/Lunts.pdf` (accessed 3/4/2009)
23. Mayer, R.: Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. American Psychologist 59(1), 14–19 (2004)
24. Merrill, M.D.: Component display theory. In: Reigeluth, C.M. (ed.) Instructional-design theories and models: An overview of their current status, Erlbaum, Hillsdale (1983)
25. Merrill, M.D., Tennyson, R.: Teaching concepts: An instructional design guide, 1st edn. Educational Technology Publications, Englewood Cliffs (1977)
26. Pressey, S.L.: A simple apparatus which gives tests and scores - and teaches. School and Society 23(586), 373–376 (1926)
27. Psotka, J., Massey, L.D., Mutter, S.A.: Intelligent Tutoring Systems: Lessons Learned. Lawrence Erlbaum Associates, Hillsdale (1988)
28. Schraagen, J.M., Chipman, S.F., Shalin, V.L.: Cognitive Task Analysis. Erlbaum, Mahwah (2000)
29. Smallwood, R.D.: A decision structure for teaching machines. MIT Press, Cambridge (1962)
30. Snow, R., Federico, P., Montague, W.: Aptitude, Learning, and Instruction, vol. 1 & 2. Erlbaum, Hillsdale (1980)
31. Stevens, A.L., Collins, A.M.: The goal structure of a Socratic tutor. In: Proceedings of the 1977 annual conference of the Association for Computing Machinery, pp. 256–263 (1977)
32. Suppes, P., Fletcher, J.D., Zanotti, M.: Models of individual trajectories in computer-assisted instruction for deaf students. Journal of Educational Psychology 68, 117–127 (1976)
33. Tobias, S.: Another look at research on the adaptation of instruction to student characteristics. Educational Psychologist 24, 213–227, Psychology 35, 389–390 (1989)
34. Tobias, S.: Interest, prior knowledge, and learning. Review of Educational Research 64, 37–54 (1994)
35. Tobias, S.: There is no such thing as "Learning styles". The CADDI Newsletter, IV(2) (2001), `http://www.eppic.biz/resources/res_newsletters_files/caddi/Vol%204%20Iss%202%20-%202001%20Summer%20Newsletter.pdf` (downloaded 3/3/2009)
36. Tobias, S.: Extending Snow's conceptions of aptitudes. Contemporary Psychology 48, 277–279 (2003)
37. Tobias, S., Fletcher, J.D. (eds.): Training and Retraining: A Handbook for Business, Industry, Government, and the Military. Macmillan Gale Group, New York (2000)
38. Weiss, D.J.: C.A.T. Central: A global resource for computerized adaptive testing (2009), `http://www.psych.umn.edu/psylabs/catcentral/` (accessed 3/4/2009)
39. Wulfeck, W.H.: A Structural Approach to Instructional Sequencing. In: Spada, H., Kempf, W. (eds.) Formalized Theories of Thinking and Learning, Huber, Bern (1976)

Assessment of Cognitive Neural Correlates for a Functional Near Infrared-Based Brain Computer Interface System

Hasan Ayaz[1], Patricia A. Shewokis[1,2], Scott Bunce[1,3], Maria Schultheis[1,4], and Banu Onaral[1]

[1] School of Biomedical Engineering, Science & Health Systems,
[2] College of Nursing and Health Professions,
[3] College of Medicine, Department of Psychiatry,
[4] College Arts and Sciences, Department of Psychology,
Drexel University, Philadelphia PA 19104, USA
{ayaz,shewokis,scott.bunce,schultheis,banu.onaral}@drexel.edu

Abstract. Functional Near Infrared Spectroscopy (fNIR) is a promising brain imaging technology that relies on optical techniques to detect changes of hemodynamic responses within the prefrontal cortex in response to sensory, motor, or cognitive activation. fNIR is safe, non-invasive, affordable, and highly portable. The objective of this study is to determine if biomarkers of neural activity generated by intentional cognitive activity, as measured by fNIR, can be used to communicate directly from the brain to a computer. A bar-size-control task based on a closed-loop system was designed and tested with 5 healthy subjects across two days. Comparisons of the average task and rest period oxygenation changes are significantly different ($p<0.01$). The average task completion time (reaching +90%) decreases with practice: day1 (mean 52.3 sec) and day2 (mean 39.1 sec). These preliminary results suggest that a closed-loop fNIR-based BCI can allow for a human-computer interaction with a mind switch task.

Keywords: Brain Computer Interface, fNIR, Near Infrared Spectroscopy.

1 Introduction

The purpose of this research is to develop a new functional Near-Infrared (fNIR) based Brain Computer Interface (BCI) to allow communication directly from the brain to a computer. In this paper, we have reported the implementation and initial results of a closed-loop fNIR based BCI system and the analysis methods that allow classification of two states (rest and task) using single channel two wavelength optical signals.

An individual's communication with the outside world can cease because of complete paralysis, locked-in syndrome, spinal cord injury or muscular dystrophy. Individuals suffering from such diseases and conditions, though conscious, may lose all voluntary muscle control and thus are often unable to communicate even their most basic wishes [1, 2]. Unlike a persistent vegetative state, in which the upper portions of

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 699–708, 2009.

the brain are damaged and the lower portions are spared, an inability to move may be caused by damage to specific portions of the lower brain and brainstem or to muscles with no damage to the upper brain. Consequently, an individual's cognitive abilities remain relatively intact [2-4].

BCI is defined as a system that translates neurophysiological signals detected from the brain to supply input to a computer or to control a device. BCI research largely targets to eliminate the need for motor movement and develop mechanisms to relay information directly from the brain to a computer which, in turn, can be used to control or communicate with outside world. In addition to their use in neuroprosthetics, noninvasive BCI systems also have potential applications for healthy individuals especially for enhancing or accelerating the learning process, or in entertainment domains such as in computer games and multimedia applications as a neurofeedback mechanism. Development of alternative communication strategies are a recognized need for clinical applications. A technique that bypasses muscles and acquires signals directly from brain would be a notable help. Moreover, this technique should be minimally intrusive, non-invasive, accessible, and safe to be used continuously.

1.1 Monitoring Brain Activity

The key element in a BCI system is monitoring brain activity. There are several available technologies that utilize different sensors or sensor configurations to collect various types of brain signals.

The most commonly studied interface to monitor brain activity noninvasively has been Electroencephalogram (EEG), due to its fine temporal resolution, portability and low cost [5-10]. Various electrode placement schemes and advanced signal processing methods have been researched for its improved and practical use in BCI applications [11]. However, these EEG based systems still have certain drawbacks. For example, the end-user has to develop a new thinking mechanism to be able to interact with the EEG based BCI system which results in lengthy training times [12]. Furthermore, non-invasive EEG recordings from portable devices are highly susceptible to noise and hence have much lower signal to noise ratio as compared to signals recorded from implanted electrodes [13]. In addition, electrode fixation is difficult and cumbersome to use in practice and for long-term use because of the need for applying gel and the restrictions on users' movements. Therefore, existing BCI systems do not yet meet the desired characteristics of an optimal BCI. In fact, they are either invasive and hence not yet completely safe for continuous use or they are non-invasive but rely on a noisy signal and require mental adaptation mechanisms.

Another potential neuroimaging modality is functional Magnetic Resonance Imaging (fMRI) which is a special type of MRI scan that measures the hemodynamic response to neural activity. Recently, this technique has been improved to be used at real-time in which output of the system could be used to give biofeedback to the subject, thus creating a closed-loop system. It has been shown using real-time functional magnetic resonance imaging (rt-fMRI) that subjects can voluntarily change activation/oxygenation levels of certain brain regions [14-22]. This technique is non-invasive and allows detecting signals anywhere in the brain, and thus provides more flexibility for the BCI mental task. However, the downside is that participants have to be scanned in large and expensive MRI machines and thus may not be practical for daily and long-term use.

In order to partially overcome the problems of existing BCI and provide an alternative communication mechanism for individuals with locked-in syndromes, we propose to use continuous wave fNIR as a new functional neuroimaging modality for Brain Computer Interface. In the next section, we will briefly discuss the fundamentals of fNIR, types of fNIR instrumentation and other BCI studies that have utilized fNIR.

2 fNIR Spectroscopy

fNIR is a multi-wavelength optical spectroscopy technique introduced as a non-invasive brain activity monitoring modality [23-27]. fNIR can assess temporal progression of brain activity, through the measurement of hemodynamic changes within reasonable spatial resolution. Neuronal activity is determined with respect to the changes in oxygenation since variation in cerebral hemodynamics are related to functional brain activity through a mechanism which is known as neurovascular coupling [26]. fNIR is not only non-invasive, safe, affordable and portable [28, 29], it also provides a balance between temporal and spatial resolution which makes fNIR a viable option for in-the field neuroimaging.

2.1 Light Tissue Interaction

Typically, an optical apparatus for fNIR Spectroscopy consists of at least one light source and a light detector that receives light after it has interacted with the tissue. Photons that enter tissue undergo two different types of interaction: absorption and scattering [30]. Whereas most biological tissues (including water) are relatively transparent to light in the near infrared range between 700 to 900 nm, hemoglobin is a strong absorber of lightwaves in this range of the spectrum.

Two chromophores, oxy- and deoxy-Hb, are strongly linked to tissue oxygenation and metabolism [26]. Fortuitously, the absorption spectra of oxy- and deoxy-Hb remain significantly different from each other allowing spectroscopic separation of these compounds to be possible by using only a few sample wavelengths. Once the photons are introduced into the human head, they are either scattered by extra- and intracellular boundaries of different layers of the head (skin, skull, cerebrospinal fluid, brain, etc.) or absorbed mainly by oxy- and deoxy-Hb. If a photodetector is placed on the skin surface at a certain distance from the light source, it can collect the photons that are scattered and thus have traveled along a “banana shaped path” from the source to the detector [23, 25, 26].

2.2 Types of fNIR Systems

A wide variety of both commercial and custom-built fNIR instruments are currently in use. There are three distinct types of fNIR spectroscopy implementations; time-resolved (TR), frequency domain (FD) and continuous wave (CW) systems, each with its own strengths and limitations. TR and FD systems provide information on shifts in both phase and amplitude of the light and are necessary for more precise quantification of fNIR signals. Lasers are used as light sources and fiber optic light guides are utilized in sensors. CW systems apply either continuous or a slow-pulsed light to tissue and measure the attenuation of amplitude of the incident light. These systems utilize less sophisticated detectors than TR and FD systems, and, therefore, they cannot determine the pathlength the photons have traveled. As such, CW systems provide only a measure of the relative change in light intensity Although CW systems provide somewhat less information than TR and FD systems, this tradeoff results in the capacity to design more compact and inexpensive hardware, making it advantageous for real-life applications [31]. CW system can use Light-Emitting-Diode (LED), instead of Laser, as light sources and do not necessarily require fiber optics in sensors, making them less expensive and more comfortable to wear for longer periods of time.

2.3 fNIR in BCI Research

There is recent evidence indicating that fNIR can be used for the assessment of attention [32] and cognitive task loads [33]. Recently, the suitability of optical methods for BCI has been investigated by acquiring signals from the motor cortex using motor imagery tasks [12, 13, 34, 35] and by acquiring signals from the frontal cortex by mental arithmetic [36] and cognitive workload [37-39] tasks. Taken together, the results of these studies have focused on offline analysis and use either FD-fNIR or laser with fiber optics. The overall aim is to build a CW-fNIR based BCI system that will be operated by the volitional activation of the prefrontal cortex assisted by neural biofeedback. As a first step, we have investigated the potential of fNIR in discriminating cognitive activity levels based on different tasks. Our results suggest that with a CW-fNIR system, we can detect increased oxygenation within the frontal lobe with increased cognitive task load [38]. In this study, we have investigated a closed-loop feedback regulated CW-fNIR based system.

3 Materials and Methods

3.1 Drexel fNIR System

The CW-fNIR system used in this study has a flexible sensor pad that contains 4 LED light sources with built in peak wavelengths at 730, 805, 850 nm and 10 detectors designed to sample cortical areas underlying the forehead (See Fig. 1). With a fixed source-detector separation of 2.5 cm, this configuration generates a total of 16 measurement channels per wavelength. The sampling rate of the system is 2Hz [32, 40, 41].

3.2 Experiment Setup

The experimental setup is composed of a Protocol-Computer, a Data-Acquisition computer and the Drexel fNIR system parts as described in Fig. 1. The fNIR sensor is positioned on the subject that is sitting in front of the Protocol Computer as shown in Fig. 2.

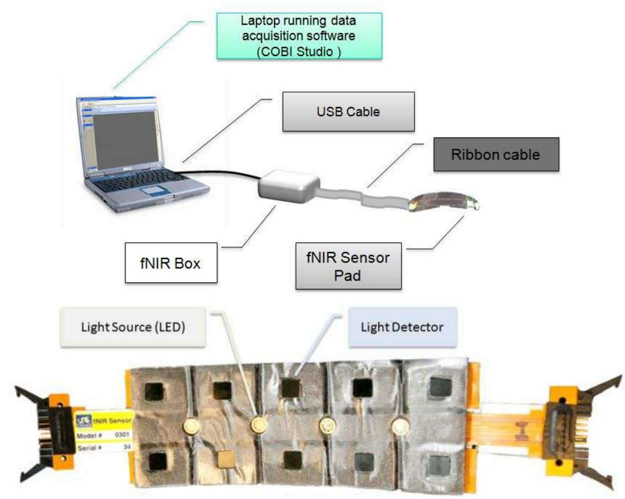


Fig. 1. Drexel fNIR System Parts

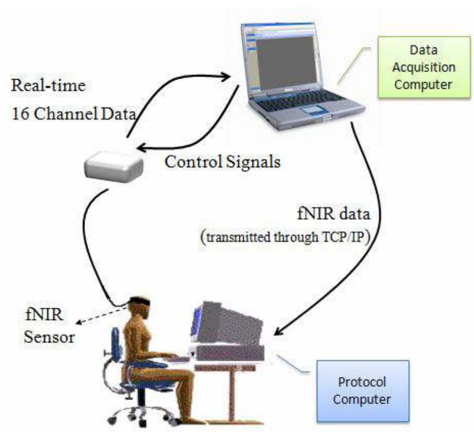


Fig. 2. Experiment Setup

Information flow starts from at the fNIR Sensor and through the control box, reaches the Data-Acquisition Computer. COBI Studio software [41] collects raw fNIR signals for 16 channels and 2 wavelengths and transmit them through Ethernet or wireless network (via TCP/IP) to the Protocol Computer. The BCI Client software on the Protocol-Computer receives the raw fNIR signals, calculates the oxygenation changes at run-time using modified Beer Lambert Law and modifies the visual feedback which in turn changes the fNIR signals at sensor; thus completing the closed loop.

3.3 Participants

Five healthy right-handed subjects (4 males, 1 female) with no neurological or psychiatric history (ages between 24 to 27years) voluntarily participated in the two-day study. Handedness was assessed by the Edinburg Handedness Inventory [42]. All subjects gave written informed consent approved by the institutional review board of Drexel University for the experiment.

3.4 Experiment Protocol

A computerized task, called bar-size-control was developed to control the timing, display the visual feedback and to save user input. In a single trial, subjects are first asked to rest for 20 seconds with a blank screen, after which a vertical or horizontal bar will appear (See Fig. 3.).

Fig. 3. Horizontal/vertical bar cue shown full-screen. The bar size is changed every 500 milliseconds according to the oxygenation changes of the subjects.

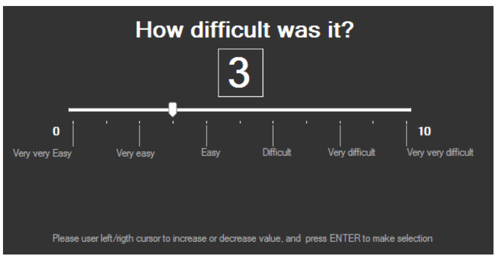


Fig. 4. Self-assessment screen is shown at the end of each trial. Subjects use left/right cursor buttons to change the value and press enter to select the value.

Initially, the bar is at 50 percent size and is mapped to the oxygenation data calculated from fNIR data that is updated at a frequency of 2Hz. The subject is asked to concentrate on the bar for up to 120 seconds. Finally, the subject is asked to rate their effort on scale from 0-10 with 0 lowest and 10 highest effort/difficulty (See Fig. 4.) [43]. The subject has 30 seconds to complete this effort rating activity. Each trial lasts a maximum of 170 seconds.

3.5 Signal Analysis

There are two types of signal processing in this study. The first one is online processing, that is done during the experiment, and the second one is offline processing that is completed after the experiment to analyze the data. Both online and offline analyses include calculation of oxygenation changes from raw data using the following steps [31, 32].

The raw optical intensity values in two wavelengths (730nm and 850nm) are transmitted and recorded by the fNIR system for all subjects. The physiologically irrelevant data (such as respiration and heart pulsation effects) and equipment noise, and so forth is first eliminated from the raw fNIR measurements by using a low-pass filter with a cut-off frequency of 0.14Hz.

The online processing further involves calculating the visual cue size based on the oxygenation changes during the experiment. Size of the vertical or horizontal bar is modeled as a linear transformation of the oxygenation changes of channel 6 that corresponds to a voxel location close to Fp1 in the international 10-20 system. Bar(t) is the bar size as a function of time t, where t_n is the time when the bar task started. BaseOxy for channel 6 and time t_n is calculated by the moving average of the last k oxygenation change values for the same channel multiplied by the constant α which is the difficulty parameter. α= 1.5 was used for all subjects.

$$Bar(t) = \left(Oxy_6(t) - BaseOxy_6^{t_n}\right) / Range_6^{t_n} * Width. \tag{1}$$

$$BaseOxy_6^{t_n} = \left(1+\alpha\right) / k \sum_{i=1}^{k} Oxy_6(n-i) . \tag{2}$$

$$Range_6^{t_n} = 2\,\alpha\; BaseOxy_6^{t_n} \tag{3}$$

For the offline processing blocks for rest and task conditions were identified for day1 and day2 of each subject. Averages of oxygenation changes in rest and task performing blocks where compared with a-repated measures ANOVA model. Furthermore, select non-parametric classification algorithms and their success rates on the available data have been applied. These techniques enable classifying a set of observations into predefined classes which in our case are task performing or resting conditions. To classify the blocks with a linear and quadratic discriminant algorithm a subset of data is used as training set. k-Nearest neighbor search (k-NN) and naive Bayes classifier (MATLAB 2008a, MathWorks Inc.) were used with day1 as training and day2 as sample, and also, half of day2 as training and the rest of day2 as sample.

4 Results and Discussion

For the computerized task, a bar was chosen for its simplicity and familiarity to all computer users. Experiments are ongoing. Comparisons of the means for task and rest period oxygenation changes are significantly different ($p<0.01$). The average task completion time (reaching +90%) decreases with practice: day1 (mean 52.3 sec) and day2 (mean 39.1 sec) across all subjects. This suggests learning and adaptation is in process.

During offline processing, blocks (rest and task periods within days) are classified with the following non-parametric algorithms: k-Nearest Neighborhood and naïve Bayes classifier. For the classification the first 16.5 seconds of each block is used. First, the algorithms are trained with the Day1 task and rest periods block data and asked to identify Day2 blocks whether they are task and rest. This was done for each subject individually and also for all subjects. The results are listed in column A in Table 1.. Next the same analysis is done with a different training set, instead of Day1, the first half of the Day2 data (task and rest periods) was used. Thus, condition B has half of the training and sample size of the previous condition. Correct classification success rates are listed in column B of Table 1.. Algorithms were unbiased and did not include the behavioral performance score or the self reported performance score. The success rate of algorithms varies between subjects suggesting that some subjects are better at using the closed loop system than other subjects. Also, column B indicates a lower success rate which in turn might be related to a lower training set than column A. Overall classification rates suggests a pattern across subjects and using the training data from all subjects provides a better chance of correct classification than individual subjects classification.

Table 1. Classification algorithm performances as percentage of correct classification in two conditions: A and B. The first condition A has training set Day1 and sample set as Day2. Second condition has training set as the first half of Day2 and sample set as second half of Day2.

	A				B			
	kNN		Bayes		kNN		Bayes	
	Rest	Task	Rest	Task	Rest	Task	Rest	Task
Subj1	100	100	100	100	100	90.9	100	90.91
Subj2	100	100	100	100	93.75	68.75	87.5	68.75
Subj3	100	80	70	55	72.72	63.64	90.1	54.55
Subj4	75	70	100	60	63.64	100	90.9	100
Subj5	80	56.67	100	100	75	68.75	87.5	75
Overall	93.3	72.5	100	95	77.1	73.77	86.89	57.37

5 Conclusion

In this study, we have reported the implementation and initial results of a closed-loop fNIR based BCI system along with the analysis methods that allows classification of two states (rest and task) using only fNIR signals. This system can be used for binary selection with volitional activation of the prefrontal cortex. Further experiments are pending to study and improve the use of algorithms for online classification.

References

1. Heiman-Patterson, T.D., Miller, R.G.: NIPPV: a treatment for ALS whose time has come. Neurology, 2006 67(5), 736–737 (2006)
2. Verma, A., Bradley, W.G.: Atypical motor neuron disease and related motor syndromes. Semin Neurol. 21(2), 177–187 (2001)
3. Abrahams, S., Leigh, P.N., Goldstein, L.H.: Cognitive change in ALS: a prospective study. Neurology, 2005 64(7), 1222–1226 (2005)
4. Neary, D., Snowden, J.S., Mann, D.M.: Cognitive change in motor neurone disease/amyotrophic lateral sclerosis (MND/ALS). J. Neurol. Sci. 180(1-2), 15–20 (2000)

5. Wolpaw, J.R., et al.: Brain-computer interfaces for communication and control. Clin. Neurophysiol. 113(6), 767–791 (2002)
6. Wolpaw, J.R., McFarland, D.J.: Multichannel EEG-based brain-computer communication. Electroencephalogr Clin. Neurophysiol. 90(6), 444–449 (1994)
7. Kubler, A., et al.: Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface. Neurology, 2005 64(10), 1775–1777 (2005)
8. Schalk, G., et al.: BCI2000: a general-purpose brain-computer interface (BCI) system. IEEE Trans Biomed. Eng. 51(6), 1034–1043 (2004)
9. Birbaumer, N., et al.: The thought translation device (TTD) for completely paralyzed patients. IEEE Trans Rehabil. Eng. 8(2), 190–193 (2000)
10. Birbaumer, N., et al.: A spelling device for the paralysed. Nature, 1999 398(6725), 297–298 (1999)
11. Wolpaw, J.R.: Brain-computer interfaces as new brain output pathways. J. Physiol. 579(Pt 3), 613–619 (2007)
12. Coyle, S., et al.: On the suitability of near-infrared (NIR) systems for next-generation brain-computer interfaces. Physiol. Meas 25(4), 815–822 (2004)
13. Sitaram, R., et al.: Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain-computer interface. NeuroImage 34(4), 1416–1427 (2007)
14. Caria, A., et al.: Regulation of anterior insular cortex activity using real-time fMRI. Neuroimage 35(3), 1238–1246 (2007)
15. De Charms, R.C.: Reading and controlling human brain activation using real-time functional magnetic resonance imaging. Trends Cogn. Sci. 11(11), 473–481 (2007)
16. de Charms, R.C., et al.: Learned regulation of spatially localized brain activation using real-time fMRI. Neuroimage 21(1), 436–443 (2004)
17. de Charms, R.C., et al.: Control over brain activation and pain learned by using real-time functional MRI. Proc. Natl. Acad. Sci. USA 102(51), 18626–18631 (2005)
18. Hinterberger, T., et al.: Neuronal mechanisms underlying control of a brain-computer interface. Eur. J. Neurosci. 21(11), 3169–3181 (2005)
19. Hinterberger, T., et al.: An EEG-driven brain-computer interface combined with functional magnetic resonance imaging (fMRI). IEEE Trans Biomed. Eng. 51(6), 971–974 (2004)
20. Weiskopf, N., et al.: Principles of a brain-computer interface (BCI) based on real-time functional magnetic resonance imaging (fMRI). IEEE Trans. Biomed. Eng. 51(6), 966–970 (2004)
21. Weiskopf, N., et al.: Real-time functional magnetic resonance imaging: methods and applications. Magn Reson Imaging 25(6), 989–1003 (2007)
22. Weiskopf, N., et al.: Physiological self-regulation of regional brain activity using real-time functional magnetic resonance imaging (fMRI): methodology and exemplary data. Neuroimage 19(3), 577–586 (2003)
23. Chance, B., et al.: Optical investigations of physiology: a study of intrinsic and extrinsic biomedical contrast. Philos. Trans R Soc. Lond. B Biol. Sci. 352(1354), 707–716 (1997)
24. Chance, B., et al.: Cognition-activated low-frequency modulation of light absorption in human brain. Proc. Natl. Acad. Sci. USA 90(8), 3770–3774 (1993)
25. Hoshi, Y., Tamura, M.: Near-infrared optical detection of sequential brain activation in the prefrontal cortex during mental tasks. Neuroimage 5(4 Pt 1), 292–297 (1997)
26. Villringer, A., Chance, B.: Non-invasive optical spectroscopy and imaging of human brain function. Trends Neurosci. 20(10), 435–442 (1997)

27. Villringer, A., et al.: Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults. Neurosci. Lett. 154(1-2), 101–104 (1993)
28. Bozkurt, A., et al.: A portable near infrared spectroscopy system for bedside monitoring of newborn brain. Biomed. Eng. Online 4(1), 29 (2005)
29. Bozkurt, A., Onaral, B.: Safety assessment of near infrared light emitting diodes for diffuse optical measurements. Biomed. Eng. Online 3(1), 9 (2004)
30. Strangman, G., Boas, D.A., Sutton, J.P.: Non-invasive neuroimaging using near-infrared light. Biol. Psychiatry 52(7), 679–693 (2002)
31. Bunce, S.C., et al.: Functional near-infrared spectroscopy: An Emerging Neuroimaging Modality. IEEE Eng Med. Biol. Mag. 25(4), 54–62 (2006)
32. Izzetoglu, M., et al.: Functional near-infrared neuroimaging. IEEE Trans Neural Syst. Rehabil. Eng. 13(2), 153–159 (2005)
33. Izzetoglu, K., et al.: Functional Optical Brain Imaging Using Near-Infrared During Cognitive Tasks. International Journal of Human-Computer Interaction 17(2), 211–227 (2004)
34. Coyle, S., Ward, T.E., Markham, C.M.: Brain-computer interface using a simplified functional near-infrared spectroscopy system. J. Neural. Eng. 4(3), 219–226 (2007)
35. Matthews, F., et al.: Hemodynamics for Brain-Computer Interfaces. Signal Processing Magazine 25(1), 87–94 (2008)
36. Bauernfeind, G., et al.: Development, set-up and first results for a one-channel near-infrared spectroscopy system. Biomed. Tech (Berl) 53(1), 36–43 (2008)
37. Naito, M., et al.: A Communication Means for Totally Locked-in ALS Patients Based on Changes in Cerebral Blood Volume Measured with Near-Infrared Light. IEICE Trans Inf. Syst. E90-D(7), 1028–1037 (2007)
38. Ayaz, H., et al.: Detecting cognitive activity related hemodynamic signal for brain computer interface using functional near infrared spectroscopy. In: 3rd IEEE/EMBS Conference on Neural Engineering, pp. 342–345 (2007)
39. Luu, S., Chau, T.: Decoding subjective preference from single-trial near-infrared spectroscopy signals. J. Neural Eng. 6(1), 16003 (2009)
40. Ayaz, H., et al.: Registering fNIR data to brain surface image using MRI templates. Conf. Proc. IEEE Eng. Med. Biol. Soc. 1, 2671–2674 (2006)
41. Ayaz, H., Onaral, B.: Analytical software and stimulus-presentation platform to utilize, visualize and analyze near-infrared spectroscopy measures, p. xiii, 103 leaves. Drexel University, Philadelphia (2005)
42. Oldfield, R.C.: The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9(1), 97–113 (1971)
43. Paas, F.G.W.C., Van Merriënboer, J.J.G.: The Efficiency of Instructional Conditions: An Approach to Combine Mental Effort and Performance Measures. Human Factors: The Journal of the Human Factors and Ergonomics Society 35, 737–743 (1993)

Systems and Strategies for Accessing the Information Content of fNIRS Imaging in Support of Noninvasive BCI Applications

Randall L. Barbour[1,2], Harry L. Graber[1,2], Yong Xu[1,2], Yaling Pei[2], Glenn R. Wylie[3], Gerald T. Voelbel[3], John DeLuca[3], and Andrei V. Medvedev[4]

[1] Department of Pathology, SUNY Downstate Medical Center, Brooklyn, NY, 11203
[2] NIRx Medical Technologies, Glen Head, NY, 11545
[3] Kessler Foundation Research Center, W. Orange, NJ, 07101
[4] Center for Functional and Molecular Imaging, Georgetown University Medical Center, Washington, DC, 20057
{randall.barbour,harry.graber,yong.xu}@downstate.edu, ypeinirx@gmail.com, {gwylie, gvoelbel, jdeluca}@kesslerfoundation.net, am236@georgetown.edu

Abstract. An essential component for a practical noninvasive brain-computer interface (BCI) system is data recording technology that can access the information-processing activity of the brain with high fidelity and throughput. Functional near-infrared spectroscopic (fNIRS) imaging is a methodology that shows promise in meeting this need, having a demonstrated sensitivity to both the slow hemodynamic response that follows neuroactivation and to the lower amplitude fast optical response that is considered a direct correlate of neuroactivation. In this report we summarize the technology integration strategy we have developed that permits detection of both signal types with a single measuring platform, and present results that document the ability to detect these data types transcranially in response to two different visual paradigms. Also emphasized is the effectiveness of different data analysis approaches that serve to isolate signals of interest. The findings support the practical utility of NIRS-based imaging methods for development of BCI applications.

Keywords: Diffuse Optical Tomography, fNIRS imaging, fast signal, combinatorial Hb States, Neuroactivation, Visual Stimulus, NIRS Technology.

1 Introduction

There are many examples where the detailing of the internal properties of otherwise opaque materials has significant value. In the late nineteen eighties, SUNY investigators first recognized that even in the limit where the penetrating energy is diffusely scattered, useful images of the internal properties of these materials was nevertheless possible [1,2]. A principal application area considered at that time for this type of imaging was the use of near infrared (NIR) light to study the optical properties of tissue [3]. Documented was the ability to generate 3D tomographic images of diffusing

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 709–718, 2009.

media whose dimensions have clinical interest based on physical models of light transport [4,5]. These preliminary findings have gone on to spur the development of a new investigative field known by the name of Diffuse Optical Tomography (DOT), or, alternately, NIRS imaging.

In the ensuing years, much effort has been directed to delineating the cost/performance trade-offs of different sensing strategies for data collection, techniques applied to forming tomographic images of diffusely scattering media, and different methods for extracting useful information from these images.

Advantages of NIRS studies include good tissue penetration, exceptional sensitivity to the hemoglobin (Hb) signal, and recovery of 3D images having a spatial resolution on the order of 1 cm. Other favorable attributes include excellent temporal resolution (msec-sec range), information about all Hb components, ease of use in different environments (including freely moving subjects), and low system cost.

Practical imaging system development is essentially an optimization problem whose limits are defined by application needs and cost/performance constraints. Typical parameters include details of the sensor array, acquisition speed and sensitivity, system control and calibration, validating phantoms, and, increasingly, access to sophisticated computing environments that support reliable feature extraction. Below we briefly outline system design strategies our group has implemented to meet these various needs, and follow this with results from experimental studies that document the capability of these designs to explore different elements of the response to neuro-activation.

1.1 Light Sensing Strategies

Brain function in the adult can be usefully probed by NIRS imaging techniques to a maximum depth of approximately 3 cm. Separation of superficial from deeply lying structures requires sampling of backreflected light using fiber-coupled sensors positioned both near and far from any source. Because fast data collection is needed to capture dynamic phenomena (*e.g.*, cerebral response to stimuli), our approach has been to employ dense sensor arrays that also have large a dynamic range of measurement. Rapid scanning is achieved by using a fast optical switch that can be operated to support all or only one of the available illumination sites. The latter arrangement allows for parallel sampling of the entire sensor array (currently up to 128 sensors) at speeds of 70-140 Hz, depending on type of signal handling circuitry. Measurements are performed using frequency-encoding techniques with homodyne detection in the audio-frequency range to allow for separate detection of light intensities from multiple illuminating wavelengths [6-8]. Tomography studies typically are conducted using a time-multiplexed, multi-site illumination approach wherein the full array is read for each illuminating site [1,3,4,6-8]. Currently feasible are illumination/sensing approaches that support sampling from four arrays in parallel, each supporting a 32S × 32D array and up to six wavelengths (4,096 illumination-detection pairs per wavelength per image frame). Such configurations can allow for tomographic imaging of approximately half the surface area of the cranium. Full head coverage can be achieved using more sparsely spaced sensor arrays. This reduces the tomography capability to a surface mapping technique known in the scientific literature as Optical Topography [9,10]. By achieving spatial separation of light signals in three dimensions, the tomography method can be expected to yield findings with greater specificity.

1.2 Sensor Head Design

The presence of hair can be an important consideration in sensor head design. Dark hair can be strongly attenuating and in such cases, to enable good fiber coupling to the scalp, careful displacement of hair is needed to achieve good signal quality. We have implemented two different design solutions. One employs open scaffolding that allows attachment of arcs that serve to mechanically support spring-loaded optical fibers. The other is a head-shaped silicone membrane that supports placement of fibers within a nearly regular array. In our experience the former is best suited for subjects with dense hair, the latter where the expected impact of hair is less important.

1.3 Anatomical Mapping

In many instances, an important object of study is to map information gained from the imaging studies to the underlying anatomy. This requires knowledge of individual head shape, of the position of the sensor array with respect to this head shape, and specification of an appropriate atlas. Currently a variety of surface-rendering tools are commercially available that have modest complexity and cost. Using methods originally developed to map EEG findings [11], we have adopted these tools to allow for mapping of NIRS image findings.

The usual case for NIRS imaging, wherein mapping of *tomographic* findings to an atlas is desired, is more complex than is typical of EEG. A key component is the need for library files that support computation of tomographic images based on a wide range of possible sensor configurations. Our approach has been to introduce a GUI that allows for easy specification of selected array configurations. The considered files are themselves based on tessellations of a segmented 3D MRI map of an adult head. Fiducial measures, along with use of affine interpolation methods, allow for accurate mapping of the sensor array to this selected atlas. Mapping to other atlases, including the individual's MRI map, is also available.

1.4 Data Analysis

Many approaches used for analysis of NIRS data for neuroimaging studies are analogues of methods developed for fMRI. Useful endpoints fall into three classes: studies on resting states, localization of activated regions, and identification of regions that are functionally connected. Because of the strong dependence of signal quality on optode separation, the quality of data across the sensor array can vary greatly. This presents the need for preprocessing schemes wherein channels having poor signal quality can be excluded from subsequent analysis [7,8].

Preprocessing is followed by use of efficient 3D image reconstruction methods that are insensitive to the usual uncertainties of experiments [12-14]. While computationally efficient, these methods tend to produce images whose accuracy and resolution can be improved using more computationally intensive techniques. The latter methods, however, have severe practical limitations when applied to image time-series studies. To this end, we have implemented alternative image correction methods that have good performance and efficiency [15-20].

1.4.1 Signal Separation Methods

Many measures from intact systems constitute a complex mixture of information over space and time. In the case of NIRS, information is convolved spatially, on a macroscopic scale, because of scattering, and temporally because of coincident phenomenology affecting different elements of the vascular tree. Compared to topographic imaging methods, image reconstruction using model-based techniques provides an objective basis for effectively reducing the blurred paths of light in tissue caused by scattering [1-5,12-20].

Among the temporal decomposition methods are techniques that can provide for isolation of signals that are uncorrelated and independent [21,22]. These methods have found favor in the functional neuroimaging community because many of the applied stimulus paradigms produce responses that largely meet these criteria. Nevertheless, because biological systems tend to operate in ways less favorable to simplifying mathematics, in many instances strict interpretation of the deconvolved time series can prove difficult. Regardless, when applied with care, these methods can prove useful and, as shown later, we have adopted one class of ICA methods to isolate the fast optical signal.

Table 1. Definitions of discrete states used to characterize hemodynamic responses. Plus sign (+) denotes an instantaneous Hb level greater than the temporal mean value; minus sign (-) denotes an instantaneous level less than the temporal mean value.

	State 1	State 2	State 3	State 4	State 5	State 6
Hb_{oxy}	-	-	-	+	+	+
Hb_{deoxy}	-	+	+	+	-	-
Hb_{total}	-	-	+	+	+	-
	Balanced	Uncompensated O_2 debt	Compensated O_2 debt	Balanced	Uncompensated O_2 excess	Compensated O_2 excess

1.4.2 Separation of Correlated Hemodynamic Signals

A specific data analysis strategy that we have applied to NIRS neuroimaging studies follows from the consideration that while there can be a time lag between an O_2 demand-linked "cause" and the subsequent blood-delivery "effect," still it is reasonable to associate different combinations of Hb levels with different conditions of balance or imbalance between the utilization and supply of O_2. For example, if Hb_{deoxy} is elevated and at the same instant Hb_{oxy} and Hb_{total} levels are reduced (in all cases, compared to their time-averaged levels in a resting baseline condition), we would interpret this as indicating that the tissue is in a state of net O_2 demand that the vasculature has not (yet) responded to by increasing the inflow of oxygenated blood. In like manner we derive the complete set of discrete combinatorial states defined in Table 1, each corresponding to a different pattern of Hb-level deviations from their baseline values in accordance with a neuroactivation induced supply-demand imbalance model.

1.5 Data Analysis Examples

In the following sections we present illustrative examples of results that have been obtained by applying an ICA algorithm to one set of NIRS imaging data, and the combinatorial state analysis to another. Common to both measurements was that a visual sensory input was used to stimulate neural activations with a prescribed time course. The distinction that determines which analysis approach is the appropriate one is that in one case a single illumination site was used (~75 Hz framing rate), and time-multiplexed, multi-site illumination (~2 Hz) was used in the other. For the first data set, ICA was used to identify the event-related fast optical signal, since it was expected that this signal would be independent of other sources of spatiotemporal variance in the data. The combinatorial-state decomposition was applied to the slow hemodynamic response data collected in the second experiment, where independence among the Hb components was not expected.

2 Methods

2.1 Data Collection

For both experiments, a multi-channel continuous wave near-infrared optical tomography imager (www.nirx.net) was used to measure, at multiple positions on each participant's scalp, the intensity of backreflected NIR light at 760 nm and 830 nm wavelengths. The optodes were positioned to make contact with the scalp, with ~1-cm inter-optode spacing, using an adjustable helmet with an open scaffolding design [23].

For the fast optical-response measurement, a 15-optode (3×5) array was positioned on the left side of the forehead. The tissue was illuminated through one of the most laterally positioned optodes, and sets of intensity measurements were collected at a ~75-Hz scan rate. For the slow hemodynamic-response measurement, measurements were performed using 30 optodes, in a 3×10 array positioned symmetrically about the midline of the occipital cortex. With each optode serving as both a source and a detector, a complete scan of the array required approximately 0.5 s.

The fast optical-response study (10 right-handed participants [6 female], 18-36 years old, mean age 26.6 yr) employed a target detection task. The visual information presented to the subjects consisted of a sequence of landscape scenes, most containing no artificial objects, while a small percentage, randomly placed within the sequence, included man-made transportation vehicles. The image presentation rate (PR) was either 4 Hz or 6 Hz, with data collected from each subject at both rates . The slow hemodynamic-response study (9 right-handed participants [2 female], 22-36 years old, mean age 27.6 yr) used a reversing-checkerboard (8 Hz) visual stimulus to induce an increase in neural activity in the visual cortex. The stimulus was presented for 2 s at a time, a total of 120 times, with a randomly varying time interval between presentations.

2.2 Data Analysis

For the fast optical-response experiment, data corresponding to each of the four wavelength-PR combinations were processed separately. For each combination, the data were frequency-filtered with a 2-30 Hz passband. Independent components were computed from each set of filtered data [24]. Any independent component (IC) that was significantly contaminated with cardiac-rhythm power, or was heavily weighted toward the superficial tissue layers, was deleted, and the remaining ICs were re-summed to produce a set of artifact-cleansed time series. Block-average responses were computed for each subject's responses to images that did and did not contain man-made objects (T [*i.e.*, target] and NT [*i.e.*, non-target], respectively). Mann-Whitney tests were performed to determine which channels and time frames showed statistically significant T responses, NT responses, or T-NT differences. Within-subject averages were computed over all channels yielding statistical significance for at least two time frames, and t-tests were performed to determine which time frames had a group-average response significantly different from zero.

In the slow hemodynamic-response case, 3D tomographic image time series Hb_{oxy} and Hb_{deoxy} were reconstructed, using the Near-infrared Analysis, Visualization and Imaging (NAVI) software package (www.nirx.net) [25]. The images were converted into ANALYZE format and exported to allow for additional processing using the AFNI image analysis suite (afni.nimh.nih.gov/afni/). Using AFNI, Hb_{total} was calculated by adding the Hb_{oxy} and Hb_{deoxy} time series. Every image-pixel time series was normalized to its temporal mean to compensate for any large differences in blood flow between individuals. The resulting scaled data were then analyzed as follows:

Area under curve. A deconvolution analysis was used to calculate an impulse response function (IRF) for the visual stimulus. The best-fitting gamma-variate function for this IRF was then determined using a nonlinear regression program [26]. This was used to calculate the event-related activation by expressing the area under the curve (AuC). To compare the AuC across participants, a t-test was performed, using a corrected voxel-level probability threshold of 0.05 ($p < 0.01$ individual voxel probability; 54-voxel cluster size). This provided a statistical test of the goodness of fit between the experimental manipulation and changes in Hb_{oxy}, Hb_{deoxy} and Hb_{total}, using standard techniques developed for the analysis of fMRI data. The correction for multiple comparisons was achieved by imposing a cluster-level threshold in addition to the voxel-level probability threshold. The cluster-level threshold, found using Monte Carlo simulations [27], was 54 contiguous voxels.

Time-fraction measures. For each voxel, the corresponding IRF was used to compute the fraction of time spent in each of the 6 combinatorial states (Table 1) over the 25 s following stimulus presentation. This resulted in a volume of the same dimensions as the reconstructed image, in which each voxel contains a number between 0 and 1 representing the fraction of time spent in one of the six states during the 25-s interval. Thus 6 volumes, one for each state, were computed for each participant. The statistical significance of the resulting time fractions was determined with t-tests comparing the observed time to a null hypothesis that was empirically determined by applying the same time-fraction analysis to images from each subject's baseline time interval.

3 Results

3.1 Measurement of Fast Optical Signal

The group-average differential T-minus-NT time course, derived from artifact-free ICs for the 830-nm, 6-Hz PR data, is plotted in Figure 1. Also included is a sketch of the measurement geometry, indicating the dimensions and coverage area of the detector array and the location of the illumination optode. Presentation of the stimulus begins at time 0 (vertical dashed line), and time frames for which the group-mean T/NT difference is statistically significant are marked with asterisks. The time delay between stimulus presentation and significant response is comparable to that typically found in electrical measurements of visual ERPs.

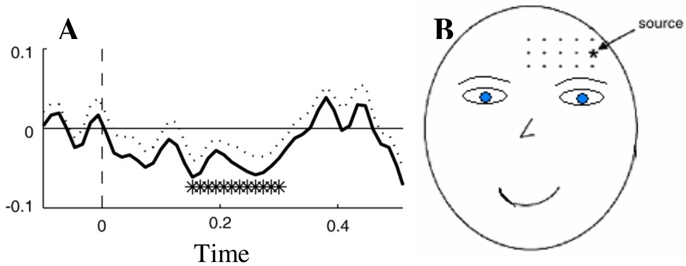


Fig. 1. Panel A: Normalized group-mean differential responses (target minus non-target) for the 830-nm, 6-Hz PR data. Dotted lines show standard errors for the corresponding signals at each time point; asterisks designate time bins with significant difference between targets and non-targets (t-test, $p < 0.05$). Panel B: geometry and location of the 15-optode array, with illumination site indicated.

3.2 Measurement of Slow Hemodynamic Response

Figure 2 shows two orthogonal sections through the center-of-mass of the region of activation, in response to the visual stimulus, identified by the group analysis of the participants' Hb_{oxy} AuC results. While many of the identified voxels are localized to the visual cortex, regions of activation outside of this location also were seen. The associated temporal response function (not shown) is triphasic, and its shape coincides well with the impulse response function seen in BOLD studies. Both the spatial- and temporal-domain results for Hb_{deoxy} and Hb_{total} are similar to those for Hb_{oxy}, while the size of the region identified as active is somewhat different in each case.

The time-fraction for each of the six combinatorial states was analyzed with a t-test across subjects that examined the difference between the fraction of time spent in each state during the 25-s stimulation-response periods and during the baseline time interval. The same correction for multiple comparisons was performed here as was used in the AuC analyses. The results, as shown in Figure 3, were that only State 1 and State 4 were identified as undergoing significant event-related responses across all nine subjects. Fig. 3 shows that the spatial extent of the time-fractions associated with States 1 and 4 both are substantially smaller than those identified in the AuC analysis of any one Hb component. It is also seen that the State-1 and State-4 regions

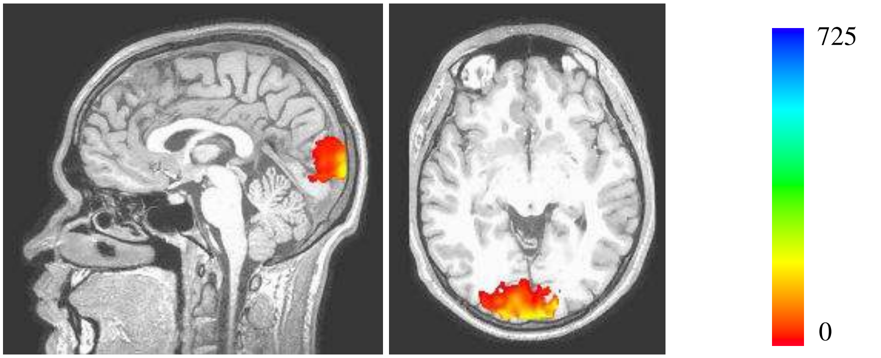


Fig. 2. Orthogonal sections through the center-of-mass of the region of activation identified by the group analysis of the participants' Hb_{oxy} AuC results

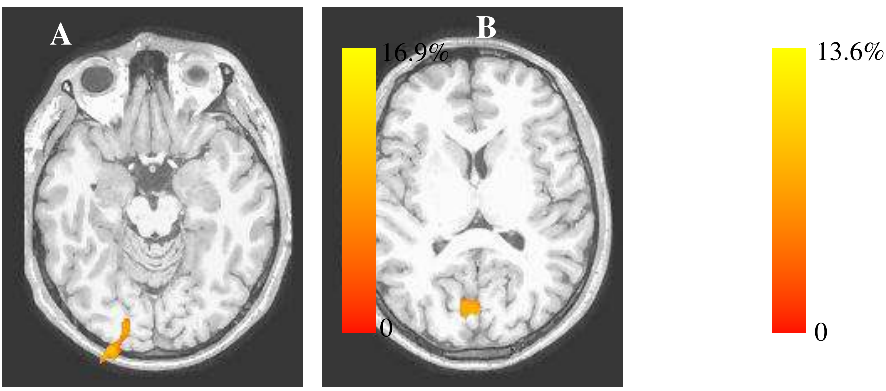


Fig. 3. Horizontal sections through the center-of-mass of the region of activation identified by the group analysis of the State 1 (Panel A) and State 4 (Panel B) time fractions

are centered in different locations, in both the Y (front-back) and Z dimensions. These results demonstrate that finer spatial resolution is achievable using analysis strategies that simultaneously consider multiple Hb components, and that differential information is associated with different combinatorial states.

4 Discussion

A research-and-development effort on functional NIRS imaging, under way since the late 1980s, seeks to identify and address all requirements for the sensing-technology associated with brain-computer interface systems. These include the ability to: 1) assess the location and magnitude of neural activity, either directly or through a surrogate parameter; 2) distinguish among different aspects of cerebral data processing (*e.g.*, sensory *vs*. cognitive); 3) examine tissue dynamics over a wide range of time scales/resolutions, with maximal freedom to specify the area being examined; 4) extract actionable, accurate information from a measurement, within a usefully brief time interval. For practicality, it also is necessary that the technology be as "transparent" as possible to the participants in a BCI application, and that it have a number of other qualities that can be classified as the "convenience" factor: ease of use, portability, ruggedness, successful performance in significantly non-ideal conditions, and low cost. The bulk of our efforts for many years went into satisfying requirements 3 and 4 above, and into clearing the transparency and convenience hurdles, and this is summarized in the Introduction. At the point that these criteria had been met, then it was appropriate to put serious effort into exploring requirements 1 and 2. The illustrative results presented here are an indication of our ability to isolate expected features of interest, strongly correlated with neural activity, from the NIRS signals generated by our technology. It is noteworthy that these encompassed measurements over two distinct cortical regions, and with very different temporal resolutions, source conditions, and data analysis strategies, but were accomplished with a single measuring platform.

Acknowledgments. This work was supported under grants 1R41NS050007-01 and 1R42NS050007-02 (RLB).

References

1. Barbour, R.L., Graber, H.L., Aronson, R., Lubowsky, J.: Model for 3-D optical imaging of tissue. In: Int. Geosci. and Remote Sensing Symp (IGARSS), pp. 1395–1399. IEEE, Piscataway (1990)
2. Aronson, R., Barbour, R.L., Lubowsky, J., Graber, H.L.: Application of transport theory to infra-red medical imaging. In: Operator Theory: Advances and Applications, vol. 51, pp. 64–75. Birkhäuser Verlag, Basel (1991)
3. Barbour, R.L., Graber, H.L., Aronson, R., Lubowsky, J.: Imaging of subsurface regions of random media by remote sensing. In: SPIE Proceedings, vol. 1431, pp. 192–203. SPIE Press, Bellingham (1991)
4. Barbour, R.L., Graber, H.L., Chang, J., Wang, Y., Aronson, R.: A perturbation approach for optical diffusion tomography using continuous-wave and time-resolved data. In: SPIE Institute of Medical Optical Tomography: Functional Imaging and Monitoring, pp. 87–120 (1993)
5. Graber, H.L., Chang, J., Aronson, R., Barbour, R.L.: A perturbation model for imaging in dense scattering media: derivation and evaluation of imaging operators. In: SPIE Institute of Medical Optical Tomography: Functional Imaging and Monitoring, pp. 121–143 (1993)
6. Schmitz, C.H., Löcker, M., Lasker, J.M., Hielscher, A.H., Barbour, R.L.: Instrumentation for fast functional optical tomography. Rev. Sci. Instrum. 73, 429–439 (2002)
7. Schmitz, C.H., Klemer, D.P., Hardin, R.E., Katz, M.S., Pei, Y., Graber, H.L., Levin, M.B., Levina, R.D., Franco, N.A., Solomon, W.B., Barbour, R.L.: Design and implementation of dynamic near-infrared optical tomographic imaging instrumentation for simultaneous dual-breast measurements. Applied Optics 44, 2140–2153 (2005)
8. Schmitz, C.H., Graber, H.L., Pei, Y., Farber, M.B., Stewart, M., Levina, R.D., Levin, M.B., Xu, Y., Barbour, R.L.: Dynamic studies of small animals with a four-color DOT imager. Rev. Sci. Instrum. 76, 094302 (2005)
9. Koizumi, H., Yamamoto, T., Maki, A., Yamashita, Y., Sato, H., Kawaguchi, H., Ichikawa, N.: Optical topography: practical problems and new applications. Applied Optics 42, 3054–3062 (2003)

10. Matsuo, K., Kato, T., Taneichi, K., Matsumoto, A., Ohtani, T., Hamamoto, T., Yamasue, H., Sakano, Y., Sasaki, T., Sadamatsu, M., Iwanami, A., Asukai, N., Kato, N.: Activation of the prefrontal cortex to trauma–related stimuli measured by near–infrared spectroscopy in posttraumatic stress disorder due to terrorism. Psychophysiology 40, 492–500 (2003)
11. Clark, C.R., Moores, K.A., Lewis, A., Weber, D.L., Fitzgibbon, S., Greenblatt, R., Brown, G., Taylor, J.: Cortical network dynamics during verbal working memory function. Int. J. Physophysiol. 42, 161–176 (2001)
12. Pei, Y., Graber, H.L., Barbour, R.L.: Influence of systematic errors in reference states on image quality and on stability of derived information for DC optical imaging. Applied Optics 40, 5755–5769 (2001)
13. Pei, Y., Graber, H.L., Barbour, R.L.: Normalized-constraint algorithm for minimizing inter-parameter crosstalk in DC optical tomography. Optics Express 9, 97–109 (2001)
14. Graber, H.L., Pei, Y., Barbour, R.L.: Imaging of spatiotemporal coincident states by DC optical tomography. IEEE Trans. Med. Imag. 21, 852–866 (2002)
15. Barbour, R.L., Graber, H.L., Xu, Y., Pei, Y., Aronson, R.: Strategies for imaging diffusing media. Transport Theory Stat. Phys. 33, 361–371 (2004)
16. Graber, H.L., Xu, Y., Pei, Y., Barbour, R.L.: Spatial deconvolution technique to improve the accuracy of reconstructed three-dimensional diffuse optical tomographic images. Applied Optics 44, 941–953 (2005)
17. Xu, Y., Graber, H.L., Pei, Y., Barbour, R.L.: Improved accuracy of reconstructed diffuse optical tomographic images by means of spatial deconvolution: two-dimensional quantitative characterization. Applied Optics 44, 2115–2139 (2005)
18. Xu, Y., Pei, Y., Graber, H.L., Barbour, R.L.: Image quality improvement via spatial deconvolution in optical tomography: Time-series imaging. J. Biomed. Optics 10, 51701 (2005)
19. Xu, Y., Graber, H.L., Barbour, R.L.: An image correction algorithm for functional 3D DOT brain imaging. Applied Optics 46, 1693–1704 (2007)
20. Graber, H.L., Xu, Y., Barbour, R.L.: An image correction scheme applied to functional DOT scattering images. Applied Optics 46, 1705–1716 (2007)
21. Stone, J.V.: Independent Component Analysis: A Tutorial Introduction. MIT Press, Cambridge (2004)
22. Stetter, M.: Exploration of Cortical Function. Kluwer Academic Publishers, Dordrecht (2002)
23. Barbour, R.L., Graber, H.L., Pei, Y., Schmitz, C.H., Di Martino, A., Castellanos, F.X.: Site-Specific Monitoring of Cerebral Vascular Hemodynamics with Dynamic Optical Tomography. Poster No. TH 252 at: Human Brain Mapping (2004)
24. Laboratory of Computer and Information Science: Adaptive Informatics Research Centre, `http://www.cis.hut.fi/projects/ica/fastica`
25. Pei, Y., Wang, Z., Xu, Y., Graber, H.L., Monteiro, R., Barbour, R.L.: NAVI: A problem solving environment (PSE) for NIRS data analysis. In: Poster No. 114 at: Fifth Inter-Institute Workshop on Optical Diagnostic Imaging from Bench to Bedside at the National Institutes of Health (2006)
26. Garavan, H., Ross, T.J., Stein, E.A.: Right hemispheric dominance of inhibitory control: an event-related functional MRI study. Proc. Natl. Acad. Sci. USA 96, 8301–8306 (1999)
27. Ward, B.D.: Simultaneous Inference for FMRI DATA, `http://afni.nimh.nih.gov/pub/dist/doc/manual/AlphaSim.pdf`

Brain-Computer Interaction

Peter Brunner[1,6] and Gerwin Schalk[1,2,3,4,5]

[1] Wadsworth Center, New York State Department of Health
[2] Department of Neurology, Albany Medical College
[3] Department of Neurosurgery, Washington University in St. Louis
[4] Department of Biomedical Engineering, Rensselaer Polytechnic Institute
[5] Department of Biomedical Sciences, State University of New York at Albany
[6] Institute for Computer Graphics and Vision, Graz University of Technology, Graz, Austria
pbrunner, schalk@wadsworth.org

Abstract. Detection and automated interpretation of attention-related or intention-related brain activity carries significant promise for many military and civilian applications. This interpretation of brain activity could provide information about a person's intended movements, imagined movements, or attentional focus, and thus could be valuable for optimizing or replacing traditional motor-based communication between a person and a computer or other output devices. We describe here the objective and preliminary results of our studies in this area.

Keywords: Brain-computer interface, BCI, Neural Engineering, Neural Prosthesis.

1 Introduction

Inquiring a soldier's directional orientation (e.g., direction of attention) is usually either impossible or at least requires motor function. This requirement is often limiting. Directly determining directional orientation from brain signals, not using muscles, would have numerous applications for military use. For example, the locus of attention and/or intended movements could be used to optimize target acquisition or identification. Brain-computer interfaces (BCIs) record signals from the brain and translate them into useful outputs. Recent studies in the rapidly growing field of BCI research provide impressive demonstrations, either with non-invasive [1-15], moderately invasive [16-20], or invasive [21-29] techniques, that BCI technology can allow people to communicate with others using brain signals alone. However, current BCI devices do not readily support large-scale deployment largely because current techniques are either not practical for use in humans [30], require extended user training [31], or function only in particular environments.

Our long-term goal is to develop BCI technologies into a range of practical and useful non-muscular communication, control, and monitoring applications. To work towards this goal, the objective of current efforts is to create a prototype of a system for communication and monitoring of orientation that uses brain signals to provide, in

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 719–723, 2009.

real time, an accurate assessment of the direction of the a person's attention, movement intention, and eye gaze.

Its achievement requires that we delineate the brain signal features associated with these variables, determine to which degree these features can be detected using non-invasive sensors, and finally create a system that can translate these features into a set of useful output functions in real time.

2 Methods

In accord with the objective outlined above, we are currently pursuing three avenues in this research. The first avenue is to delineate the brain signal features associated with the direction of attention, intention, and eye gaze. We do this by recording electrical brain signals invasively from the surface of the brain in human subjects. These subjects are asked to engage in tasks that are designed to vary relevant parameters, such the direction of attention. (These subjects are human patients that have electrode grids implanted on the surface of the brain for clinical reasons. Thus, they are not being implanted for the purpose of these research projects.) Analyses relate the brain signals to the parameters of interest described above, and thereby delineate the brain signal features that are most predictive of the particular task. We also plan to determine the relationship of the observed features across time and space to establish a mechanistic understanding of relevant cortical systems.

The second avenue of research is to determine whether these features can be detected non-invasively. To provide information that is critical for practical deployment of such a system, we plan to determine the degree to which these brain signal features and brain systems can also be detected using electrical sensors placed non-invasively on the scalp. To do this, we will use the results gathered using invasive methodologies to guide electrode placement and analyses.

The third avenue of research is to validate the use of brain signals for communication and orientation. To create a prototype of an intuitive communication and orientation system, we will design algorithms that are capable of extracting the features identified above on a single-trial basis, and incorporate these procedures into a real-time software system, called BCI2000 [32], that has been developed in our laboratory over the past decade. The resulting system will allow for a real-time assessment of the direction of the user's attention, intended movements, and eye gaze.

We expect that these efforts will provide the first prototype system that can derive these parameters in humans in real time. These efforts should also contribute fundamental neuroscientific understanding in humans.

3 Preliminary Results

Preliminary results to date provide encouraging evidence that brain signals in humans hold information (which could be extracted in real time) about the direction of attention (Fig. 1) and eye gaze.

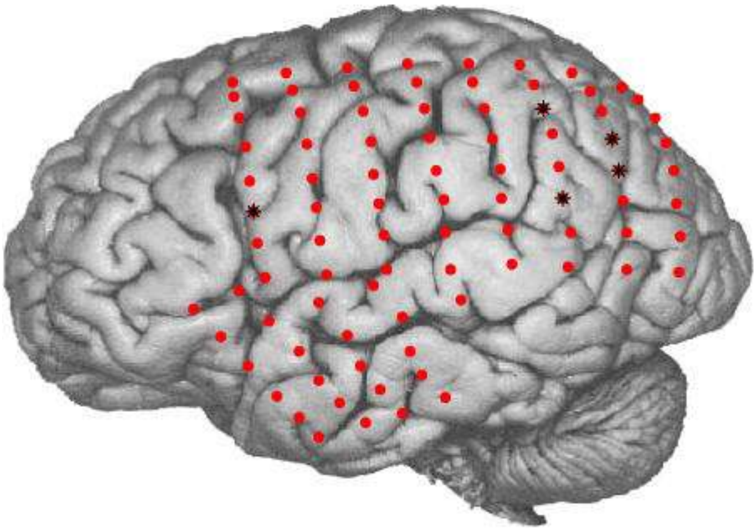


Fig. 1. Information about directional attention in one subject. Red dots indicate all electrode locations. Black stars indicate locations that hold statistically significant ($p<0.001$, Bonferroni corrected) information about whether the subject's attention is focused on the left or right hemisphere of the visual field.

4 Conclusions

Our ongoing studies are addressing the question whether it is possible to derive, in real time, signals from the brain in humans that provide information about a person's directional orientation, i.e., attention, intention, and eye gaze. Preliminary results support this hypothesis.

Acknowledgement

This work was supported by the US Army Research Office (W911NF-07-1-0415 and W911NF-08-1-0216) and the NIH/NIBIB (EB006356 and EB00856).

References

1. Farwell, L.A., Donchin, E.: Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Clinical Neurophysiology 70, 510–523 (1988)
2. Wolpaw, J.R., McFarland, D.J., Neat, G.W., Forneris, C.A.: An EEG-based brain-computer interface for cursor control. Electroencephalography and clinical neurophysiology 78, 252–259 (1991)
3. Sutter, E.E.: The Brain Response Interface: Communication Through Visually-Induced Electrical Brain Responses. J. Microcomp. App. 15, 31–45 (1992)
4. McFarland, D.J., Neat, G.W., Wolpaw, J.R.: An EEG-based method for graded cursor control. Psychobiology 21, 77–81 (1993)

5. Pfurtscheller, G., Flotzinger, D., Kalcher, J.: Brain-Computer Interface - a New Communication Device for Handicapped Persons. J. Microcomp. App. 16, 293–299 (1993)
6. Wolpaw, J.R., McFarland, D.J.: Multichannel EEG-based brain-computer communication. Clin. Neurophysiol. 90, 444–449 (1994)
7. Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, A., Perelmouter, J., Taub, E., Flor, H.: A spelling device for the paralysed. Nature 398, 297–298 (1999)
8. Pfurtscheller, G., Guger, C., Mueller, G., Krausz, G., Neuper, C.: Brain oscillations control hand orthosis in a tetraplegic. Neurosci. Lett. 292, 211–214 (2000)
9. Millan, J.R., Renkens, F., Mourino, J., Gerstner, W.: Noninvasive brain-actuated control of a mobile robot by human EEG. IEEE Trans Biomed. Eng. 51, 1026–1033 (2004)
10. Wolpaw, J.R., McFarland, D.J.: Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. Proc. Natl. Acad. Sci. U.S.A. 101, 17849–17854 (2004)
11. Kübler, A., Nijboer, F., Mellinger, J., Vaughan, T.M., Pawelzik, H., Schalk, G., McFarland, D.J., Birbaumer, N., Wolpaw, J.R.: Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface. Neurology 64, 1775–1777 (2005)
12. Blankertz, B., Dornhege, G., Krauledat, M., Muller, K.R., Kunzmann, V., Losch, F., Curio, G.: The Berlin Brain-Computer Interface: EEG-based communication without subject training. IEEE Trans Neural Syst. Rehabil. Eng. 14, 147–152 (2006)
13. Vaughan, T.M., McFarland, D.J., Schalk, G., Sarnacki, W.A., Krusienski, D.J., Sellers, E.W., Wolpaw, J.R.: The Wadsworth BCI Research and Development Program: at home with BCI. IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society 14, 229–233 (2006)
14. McFarland, D., Krusienski, D., Sarnacki, W., Wolpaw, J.: Emulation of computer mouse control with a noninvasive brain–computer interface. J. Neural Eng. 5, 101–110 (2008)
15. Muller, K.R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., Blankertz, B.: Machine learning for real-time single-trial EEG-analysis: from brain-computer interfacing to mental state monitoring. J. Neurosci. Methods 167, 82–90 (2008)
16. Wilson, J.A., Felton, E.A., Garell, P.C., Schalk, G., Williams, J.C.: ECoG factors underlying multimodal control of a brain-computer interface. IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society 14, 246–250 (2006)
17. Leuthardt, E.C., Miller, K.J., Schalk, G., Rao, R.P., Ojemann, J.G.: Electrocorticography-based brain computer interface–the Seattle experience. IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society 14, 194–198 (2006)
18. Leuthardt, E.C., Schalk, G., Wolpaw, J.R., Ojemann, J.G., Moran, D.W.: A brain-computer interface using electrocorticographic signals in humans. Journal of Neural Engineering 1, 63–71 (2004)
19. Felton, E.A., Wilson, J.A., Williams, J.C., Garell, P.C.: Electrocorticographically controlled brain-computer interfaces using motor and sensory imagery in patients with temporary subdural electrode implants. Report of four cases. J. Neurosurg. 106, 495–500 (2007)
20. Schalk, G., Miller, K., Anderson, N., Wilson, J., Smyth, M., Ojemann, J., Moran, D., Wolpaw, J., Leuthardt, E.: Two-dimensional movement control using electrocorticographic signals in humans. J. Neural Eng. 5, 75–84 (2008)
21. Velliste, M., Perel, S., Spalding, M.C., Whitford, A.S., Schwartz, A.B.: Cortical control of a prosthetic arm for self-feeding. Nature 453, 1098–1101 (2008)

22. Hochberg, L., Serruya, M., Friehs, G., Mukand, J., Saleh, Caplan, A., Branner, Chen, Penn, R., Donoghue, J.: Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature 442, 164–171 (2006)
23. Santhanam, G., Ryu, S.I., Yu, B.M., Afshar, A., Shenoy, K.V.: A high-performance brain-computer interface. Nature 442, 195–198 (2006)
24. Lebedev, M.A., Carmena, J.M., O'Doherty, J.E., Zacksenhouse, M., Henriquez, C.S., Principe, J.C., Nicolelis, M.A.: Cortical ensemble adaptation to represent velocity of an artificial actuator controlled by a brain-machine interface. J. Neurosci. 25, 4681–4693 (2005)
25. Andersen, R.A., Burdick, J.W., Musallam, S., Pesaran, B., Cham, J.G.: Cognitive neural prosthetics. Trends Cogn. Sci (Regul. Ed.) 8, 486–493 (2004)
26. Shenoy, K.V., Meeker, D., Cao, S., Kureshi, S.A., Pesaran, B., Buneo, C.A., Batista, A.P., Mitra, P.P., Burdick, J.W., Andersen, R.A.: Neural prosthetic control signals from plan activity. Neuroreport 14, 591–596 (2003)
27. Serruya, M., Hatsopoulos, N., Paninski, L., Fellows, M., Donoghue, J.: Instant neural control of a movement signal. Nature 416, 141–142 (2002)
28. Taylor, D.M., Tillery, S.I., Schwartz, A.B.: Direct cortical control of 3D neuroprosthetic devices. Science 296, 1829–1832 (2002)
29. Georgopoulos, A.P., Schwartz, A.B., Kettner, R.E.: Neuronal population coding of movement direction. Science 233, 1416–1419 (1986)
30. Shain, W., Spataro, L., Dilgen, J., Haverstick, K., Retterer, S., Isaacson, M., Saltzman, M., Turner, J.: Controlling Cellular Reactive Responses Around Neural Prosthetic Devices Using Peripheral and Local Intervention Strategies. IEEE Trans Neural Syst. Rehabil. Eng. 11, 186–188 (2003)
31. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain-computer interfaces for communication and control. Clin. Neurophysiol. 113, 767–791 (2002)
32. Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N., Wolpaw, J.R.: BCI2000: a general-purpose brain-computer interface (BCI) system. IEEE transactions on bio-medical engineering 51, 1034–1043 (2004)

P300 Based Brain Computer Interfaces: A Progress Report

Emanuel Donchin and Yael Arbel

Department of Psychology, University of South Florida, Tampa, FL
donchin@shell.cas.usf.edu, yarbel@mail.usf.edu

Abstract. Brain-Computer Interfaces (BCI) are the only means of communication available to patients who are locked-in, that is for patients who are completely paralyzed yet are fully conscious. We focus on the status of the P300-BCI first described by Farwell and Donchin (1988). This system has now been tested with several dozen ALS patients and some have been using this approach for communication at a very extensive level. More recently, we have adapted this BCI (in collaboration with the laboratory of Dr. Rajiv Dubey) to the control of a robotic arm. In this presentation we will discuss the special problems of human computer interaction that occur within the context of such a BCI. The special needs of the users forced the development of variants of this system, each with advantages and disadvantages. The general principles that can be derived from the experience we have had with this BCI will be reviewed.

Keywords: Brain Computer Interface (BCI), P300, wheelchair-mounted robotic arm (WMRA).

1 Introduction

A Brain Computer Interface (BCI) is a device that allows users to communicate with the world without utilizing voluntary muscle activity (i.e., using only the electrical activity of the brain). Several BCI programs were established with a focus on developing new augmentative communication and control technology for those with severe neuromuscular disorders. BCI systems utilize what is known about electrical brain activity to detect the message that a user has chosen to communicate. These systems rely on the finding that the brain reacts differently to different stimuli, based on the level of attention given to the stimulus and the specific processing triggered by the stimulus. Described by Farewell and Donchin in 1988 [1], the P300 based Speller is one such BCI system that relies on a brain response known as the P300, whose attributes have been studied for over four decades.

1.1 What Is a P300?

The P300, first described by Sutton, Braren, Zubin, & John (1965) [2], is one of the components of the brain's response to specific events that can be recorded from the scalp. These "event related potentials (ERPs) are manifestations of brain activities invoked in the course of information processing. The P300 reaches its maximal

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 724–731, 2009.

amplitude at least 300 ms following rare task-relevant stimuli. It is the largest at the parietal electrodes, somewhat smaller at the central electrodes and minimal at the frontal electrodes.

The P300 is elicited by rare task-relevant events and if often recorded in what has come to be called the “oddball” paradigm [3]. The “oddball” paradigm requires the participant to apply a classification rule to each of the events in a random sequence of events so that each event is classified into one of two categories, one of which is presented infrequently. The participant is required to perform a task that cannot be accomplished without the categorization of the events. As the P300 is elicited by events belonging to the rare category, its latency varies with the time required for categorizing the events. The amplitude of the P300 varies with the subjective probability and the task relevance of the eliciting events. Thus, the rarer the event, the larger the P300 it elicits.

1.2 P300 Based BCI

Two decades ago, Farwell and Donchin [1] developed a P300 based BCI that enables individuals to communicate with their environment without using any neuromuscular function. This P300 BCI speller uses an Oddball paradigm to elicit a P300 to a character that the user is choosing to communicate. The user is presented with a visual matrix of characters. The rows and columns of this matrix are flashed in a random sequence. The user focuses attention on one character to be communicated. Flashes of the row and column of the attended character are the rare events in this “oddball paradigm”, Flashes of the other rows and columns compose the frequent events. Thus, the flashes of rows and columns containing the attended character elicit a P300, while rows and columns not containing this letter do not elicit a P300. Therefore, by computing the ERPs associated with flashes of every row and column in the matrix, and detecting which row and column elicited a P300 response, the BCI system can identify in real time the character the user chose to communicate.

The size of the matrix can be varied according to individual preferences and ability. The matrix’ cells may contain letters, numbers, words, sentences, pictures and/or symbols. Depending on the user’s needs and preferences, the matrix can be as small as a 2x2 with 4 stimuli (for example, “yes”, “no”, “stop”, “more”), or as large as a 9x8 to emulate a computer keyboard. The successful use of the system does not require any training of the user. However, for optimal use, the algorithm detecting the P300 needs to be “calibrated” based on the pattern of electrical brain activity of a specific user.

1.3 Speed-Accuracy Tradeoffs

As the detection of P300 requires signal averaging, a number of trials are required by the system to correctly determine the user’s selection. The speed of the system thus depends on the number of sequences of flashes required to achieve a given level of accuracy. Traditionally, speed-accuracy tradeoff is estimated by analyzing a dataset offline to evaluate the number of events the system needed to average to achieve the desired accuracy level. However, the offline analysis does not take into account factors that are related to the user of the system (e.g., ability to sustain attention during longer trials).

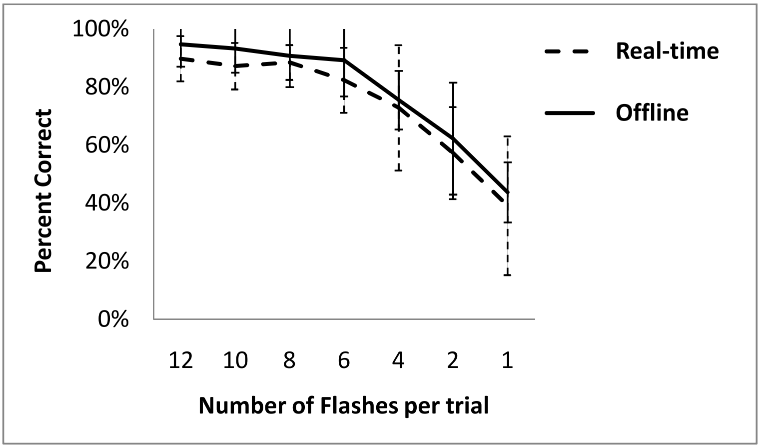


Fig. 1. Accuracy as a function of number of flashes per trial in real time and as estimated by offline analysis

We have recently examined the speed-accuracy tradeoffs of the P300 BCI speller measured in real time while participants selected characters from a 6 by 6 matrix with letters and numbers. Six young adults from the University of South Florida attended five 2-hour sessions to evaluate accuracy of spelling while manipulating the number of events (flash sequences). Accuracy was evaluated while participants spelled 50 characters under each of seven conditions: when each of the 12 rows and columns flashed twelve, ten, eight, six, four times, twice, and once. These speed accuracy data are reported in comparison to the data obtained from the offline analysis. Our results (Fig. 1) validate the effectiveness of the offline speed-accuracy estimation, although greater variability in accuracy was found in real time, particularly when a single sequence of flashes was used per character.

1.4 Adapting the BCI System for the Use of ALS Patients

As of today, most of the users of the BCI system are patients with Amyotrophic lateral sclerosis (ALS). Approximately 5,600 people in the U.S. are diagnosed with ALS each year. ALS, also called Lou Gehrig's disease, is a progressing, neurological disease that attacks the neurons responsible for controlling voluntary muscles. For the vast majority of people with ALS, their minds and thoughts are unaffected, remaining mentally sharp despite the progressive degeneration of their bodies.

With modern technology and advanced healthcare services, patients with ALS live longer. About twenty percent of people with ALS live five years or more and up to ten percent will survive more than ten years and five percent will live 20 years. As the progression of the disease is commonly rapid, and as the loss of the ability to function independently is relatively early, it is extremely important to provide these patients with a mean of performing everyday tasks even in the "locked-in" stage of the disease in which they can stay for years.

Extensive studies with ALS patients have demonstrated that the P300 BCI system can allow communication at the rate of 8 characters per minute. Since 2002, Sellers and Donchin [4] have tested the system with some 25 ALS patients at different stages of the disease in the Cognitive Psychophysiological Laboratory at the University of South Florida. A study by Sellers and Donchin [4] indicates that a P300-based BCI system can be successfully operated by patients suffering from ALS. In this study, a simplified version of the P300-speller was used. The reason for this simplification was that it was difficult for some patients to use the 6 by 6 letter matrix to spell out words. Therefore, the user focused attention to one of just four response options: "yes", "no", "pass" and "end", which were displayed and randomly flashed on a computer screen. Users were asked to either focus attention on one item, or to select the correct answer to a question asked by the experimenter. The results showed that ALS patients are able to reliably use a P300-based BCI.

Nijboer et al. (2008) [5] evaluated the efficacy of a P300 BCI speller for individuals with advanced ALS. In Phase I, six participants used a 6 x 6 matrix on 12 separate days with a mean rate of 1.2 selections/min and mean online and offline accuracies of 62% and 82%, respectively. In Phase II, four participants used either a 6x6 or a 7x7 matrix to produce novel and spontaneous statements with a mean online rate of 2.1 selections/min and online accuracy of 79%. The amplitude and latency of the P300 remained stable over 40 weeks. The results demonstrated that people who are severely disabled by ALS could communicate with the P300-based BCI and performance was stable over many months.

1.5 The P300 BCI Controls and Operates a Robotic Arm Mounted to a Wheelchair

Originally, EEG-based BCI systems were adapted to control simple functions, such as choosing letters from a screen to spell out words (e.g., [1], [6], [7], [8]), or moving a cursor on a screen. More recently, attempts have been made to adapt BCIs to steer robots (e.g., [9], [10]) and wheelchairs (e.g., [11], [12], [13], [14]), as well as to control implantable neuroprostheses [15] and robot arms [16]. Research on BCIs controlling these new devices is in a very early stage. We have recently demonstrated that the P300 BCI can be used to communicate a selected character from a 5x3 matrix to the controller of a wheelchair-mounted robotic arm (WMRA) [17] (see illustration of the communication between the BCI and the Robotic arm in Fig 4) [18]. To control the WMRA via the BCI the user is presented with a visual matrix whose rows and columns intensify randomly. Each of the symbols in the matrix corresponds to a specific direction or task command (Fig. 2). The chosen character from the BCI display is sent to the WMRA control program, which translates it into a Cartesian velocity in the proper direction and executes the algorithm to move the arm.

To test the application of the P300 BCI as a controller of the WMRA, six healthy young adults from the University of South Florida were presented with a 5x3 visual matrix with letters (see Fig. 3). Every row and column intensified for 75 ms every 50 ms. Each sequence of flashes contained 8 intensifications (5 columns and 3 rows) and lasted for 1 sec. We tested the accuracy of character selection as a function of number of sequences of flashes (number of intensifications). The letters in the BCI display (Fig. 3) corresponded with the symbol matrix of the WMRA interface (Fig. 2). In other words, the user was presented with the alphabet speller matrix, which was mapped to the robot actions. For example, the letter “B” corresponds with the arrow directing the robot to move forward.

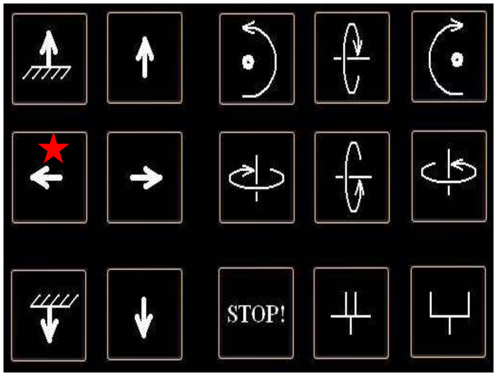


Fig. 2. Display of the Robotic arm controller

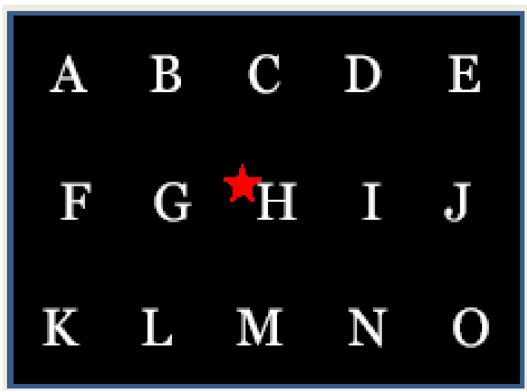


Fig. 3. A 5 X 3 display of the BCI. Each letter corresponds to a specific direction of the WMRA as seen in Fig 2.

Fig. 4 illustrates the operation of the WMRA via the BCI. In Fig. 5 is a user operating the WMRA by choosing characters from the BCI display. For safety of the user, the movement of the robotic arm was kept slow by keeping the scaling factor low.

Accuracy level was measured by comparing the character to spell with the character selected by the BCI system after it examines the recorded data in real time. Number of flash sequences may be viewed as the amount of data that were available for averaging and signal extraction. It can also be discussed in terms of speed as the more flash sequences were collected for each character, the longer the trial was before the system reached a decision. As was expected, accuracy dropped as a function of flash sequences. However, this reduction in accuracy level was minimal to moderate. When asked, participants informed the tester that they preferred the 4 and 6 sequences of flashes over the longer sequences. The common explanation was that it was easier to stay focused for shorter periods of time. Below is the accuracy data obtained when participants spelled 50 characters of each set of sequences (12, 10, 8, 6, 4, and 2). Fig. 6 shows the mean percentages correct for each sequence. Number of maximum characters per min and number of correct characters per minute are also presented.

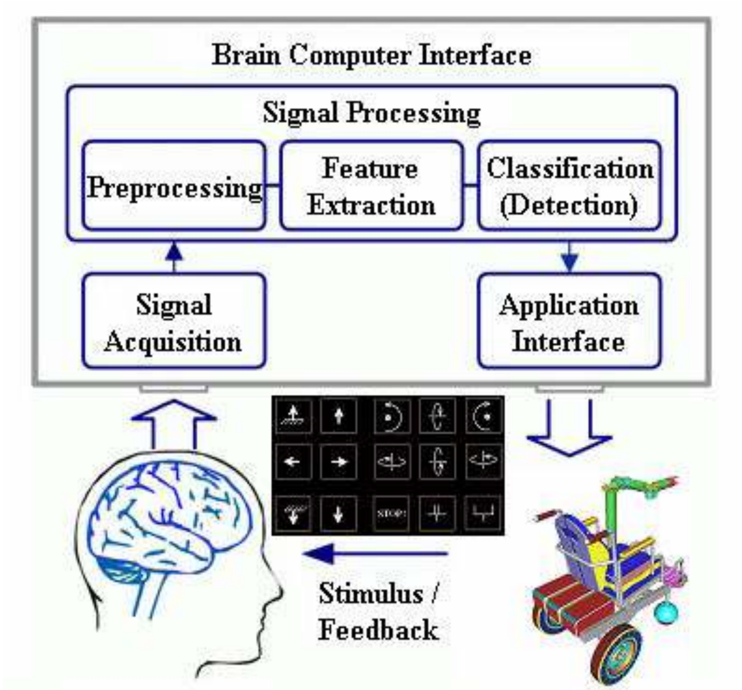


Fig. 4. An illustration of the communication between the BCI and the controller of the WMRA

Fig. 5. A user controlling the WMRA by choosing characters from the BCI display

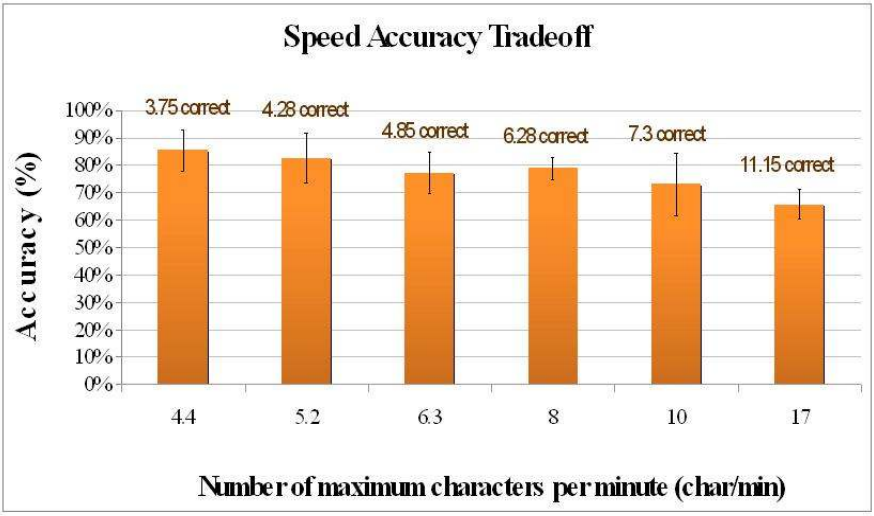


Fig. 6. Accuracy data (% correct) for each of the # of flash sequences (from left to right: 12, 10, 8, 6, 4, 2). For each bar of the # of flash sequences we provide the maximum number of characters per minute (on the bottom of each bar) and the number of correct characters per minute (on top of each bar).

There are a few potential challenges which merit consideration. The step by step manipulation of the arm is not effective in reaching our goal for a system that will be used for daily activities such as bringing a glass of water from the kitchen or opening the door. Rather than characters representing one specific movement, the display of the BCI should contain high-level commands, which can then be executed autonomously by the robot via task level planning control. The challenge is to develop a system that will be able to dynamically estimate and represent the user's intentions in relations to the changing environment, to communicate these intentions in the most efficient manner to the robotic arm which will have the intelligence to perform the task effectively and safely. More specifically, our current goals are to transform the mobile robotic arm into a task oriented system which is programmed to perform tasks in a changing environment efficiently, program the Application Module in the BCI system to represent the environment and the user's intentions effectively and in a flexible manner, and to improve the speed of task selection by evaluating alternative classification techniques with a goal of detecting the P300 using a substantially smaller number of trials than is currently required using the Stepwise Discriminant analysis in the current version of the P300 Speller.

References

1. Farwell, L., Donchin, E.: Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. Electro-encephalography and Clinical Neurophysiology, 510–523 (1988)
2. Sutton, S., Braren, M., Zublin, J., John, E.: Evoked Potential Correlates of Stimulus Uncertainty. Science 150, 1187–1188 (1965)

3. Donchin, E., Ritter, W., McCallum, C.: Cognitive psychophysiology: The endogenous components of the ERP. In: Callaway, E., Tueting, P., Koslow, S. (eds.) Brain event-related potentials in man, pp. 349–441. Academic Press, New York (1978)
4. Sellers, E.W., Donchin, E.: A P300-based brain-computer interface: Initial tests by ALS patients. Clinical Neurophysiology, 538–548 (2006)
5. Nijboer, F., Sellers, E.W., Mellinger, J., Jordon, M.A., Matuz, T., Furdea, A., Halder, S., Mochty, U., Krusienski, D.J., Vaughan, T.M., Wolpaw, J.R., Birbaumer, N., Kübler, A.: A P300-based brain-computer interface for people with amyotrophic lateral sclerosis. Clinical Neurophysiology 119(8), 1909–1916 (2008)
6. Donchin, E., Spencer, K., Wijesinghe, R.: The mental prosthesis: assessing the speed of a P300-based brain-computer interface. IEEE Trans Rehabil. Eng. 8, 174–179 (2000)
7. Hinterberger, T., Kübler, A., Kaiser, J., Neumann, N., Birbaumer, N.A.: Brain-computer interface (BCI) for the locked-in: comparison of different EEG classifications for the thought translation device. Clinical Neurophysiology 114, 416–425 (2003)
8. Obermaier, B., Müller, G.R., Pfurtscheller, G.: 'Virtual Keyboard' Controlled by Spontaneous EEG Activity. IEEE Transactions on Neural Systems and Rehabilitation Engineering 11(4), 422–426 (2003)
9. Bell, C.J., Shenoy, P., Chalodhorn, R., Rao, R.P.N.: An Image-based Brain-Computer Interface Using the P3-Response. In: Proceedings of the 3rd International IEEE EMBS Conference on Neural Engineering, pp. 318–321 (2007)
10. Hinic, V., Petriu, E.M., Whalen, T.E.: Human-Computer Symbiotic Cooperation in Robot-Sensor Networks. In: Instrumentation and Measurement Technology Conference – IMTC, pp. 1–5 (2007)
11. Rebsamen, B., Burdet, E., Guan, C., Teo, C.L., Zeng, Q., Ang, M., Laugier, C.: Controlling a wheelchair using a BCI with low information transfer rate. In: Proceedings of the 2007 IEEE International Conference on Rehabilitation Robotics, pp. 1003–1008 (2007)
12. Millan, J.D.R., Renkens, F., Mourino, J., Gerstner, W.: Noninvasive Brain-Actuated Control of a Mobile Robot by Human EEG. IEEE Transactions on Biomedical Engineering 51(6), 1026–1033 (2004)
13. Tanaka, K., Matsunaga, K., Wang, H.O.: Electroencephalogram-Based Control of an Electric Wheelchair. IEEE Transactions on Robotics 21(4), 762–766 (2005)
14. Philips, J., Del, R., Millan, J., Vanacker, G., Lew, E., Galan, F., Ferrez, P.W., Van Brussel, H., Nuttin, M.: Adaptive Shared Control of a Brain-Actuated Simulated Wheelchair. In: Proceedings of the 2007 IEEE 10th International Conference on Rehabilitation Robotics, pp. 408–414 (2007)
15. Müller-Putz, G.R., Scherer, R., Pfurtscheller, G., Rupp, R.: EEG-based neuroprosthesis control: A step towards clinical practice. Neuroscience Letters 382, 169–174 (2005)
16. Lüth, T., Ojdanic, D., Friman, O., Gräser, A.: Low level control in a semi-autonomous rehabilitation robotic system via a Brain-Computer Interface. In: Proceedings of the 2007 IEEE 10th International Conference on Rehabilitation Robotics, pp. 721–728 (2007)
17. Alqasemi, R., Dubey, R.: Maximizing Manipulation Capabilities for People with Disabilities Using a 9-DoF Wheelchair-Mounted Robotic Arm System. In: Proc. of the 2007 International Conf. on Rehabilitation Robotics, Noordwijk, Netherlands (2007)
18. Arbel, Y., Alqasemi, R., Dubey, R., Donchin, E.: Adapting the P300-Brain Computer Interface (BCI) for the control of a wheelchair-mounted robotic arm system. Psychophysiology 44, S1, S82–S83 (2007)

Goal-Oriented Control with Brain-Computer Interface

Günter Edlinger[1], Clemens Holzner[1], Christoph Groenegress[2], Christoph Guger[1], and Mel Slater[2,3]

[1] g.tec Guger Technologies OEG, Herbersteinstrasse 60, 8020 Graz,
[2] Universitat Politècnica de Catalunya, Barcelona, Spain
[3] Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain
edlinger@gtec.at

Abstract. A brain-computer interface (BCI) is a new communication channel between the human brain and a digital computer. Such systems have been designed to support disabled people for communication and environmental control. In more recent research also BCI control in combination with Virtual Environments (VE) gains more and more interest. Within this study we present experiments combining BCI systems and VE for navigation and control purposes just by thoughts. Results show that the new P300 based BCI system allows a very reliable control of the VR system. Of special importance is the possibility to select very rapidly the specific command out of many different choices. The study suggests that more than 80% of the population could use such a BCI within 5 minutes of training only. This eliminates the usage of decision trees as previously done with BCI systems.

Keywords: Brain-computer interface, virtual reality, P300 evoked potential.

1 Introduction

A Brain-Computer interface (BCI) is a new communication channel allowing subjects to interact with a computer without using any muscle activity. Such a system represents an additional output channel without relying on the brain's normal pathways of muscles or peripheral nerves [1, 2]. A BCI converts specific brain signals into control commands using pattern recognition methods. In order to properly operate a BCI, the system is firstly trained on subject specific brain activity data.

Brain-computer interface systems have been developed during the last years for people with severe disabilities to improve their quality of life. Applications of BCI systems comprise the restoration of movements, communication and environmental control [1-3]. However, recently BCI applications have been also used in different research areas e.g. in the field of virtual reality [4, 5]. There the control of and navigation in smart homes via BCI interfaces can be studied before realizing the real world smart home environment.

Non-invasive BCI systems have been successfully realized based on different brain electrical signal (electroencephalogram, EEG) phenomena:

Between 1 and 5 degrees of freedom of control have been realized up to now for slow cortical potentials [1]. Steady-state visual evoked potentials [6;7] allow mostly

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 732–740, 2009.

up to 12 different decisions and are only limited by the number of distinct frequency responses that can be analyzed in the EEG. This approach uses the fact that flickering light sources with flickering frequencies in the range of around 8-20 Hz induce brain oscillations of the same flickering frequency. Applications so far comprise e.g. robot or mobile phone control [8].

BCI systems based on induced oscillations use mostly motor imagery strategies to generate event-related de-/synchronization (ERD/ERS) in the alpha and beta frequency ranges of the EEG [5, 9] This type of BCI was realized for cursor control on computer screens, for navigation of wheelchairs or in virtual environments [5]. About 2-4 degrees of freedom for control can be realized so far. However, it remains the case that the highest information transfer rates are reached with only 2 decisions beyond which the accuracy falls dramatically.

A P300 based BCI system uses the effect that an unlikely event induces a P300 component in the EEG, i.e. a positive deflection in the EEG signal is occurring around 300 ms after the event. Such systems are suited for spelling device, because a high number of different target characters enhance the BCI communication speed [7, 10, 11]. However, recently BCI interfaces in e.g. Japanese language using up to 72 letters have also been reported [12].

In a spelling application characters or icons are ordered in rows and columns on the computer screen. There exist two different strategies to realize the P300 speller: (i) the row /column (RC) speller highlight multiple characters at once and the single character (SC) speller flashes each character individually.

Therefore a higher P300 amplitude and more reliable control can be expected with the SC flasher because it is more unlikely that the target character appears. Sellers found that a 3 x 3 matrix had higher accuracy than a 6 x 6 matrix, but a lower communication rate. With an inter-stimulus interval (ISI) of 175 ms and a 3x3 matrix Sellers achieved an accuracy of 88 % in the best case [11].

This study is divided into two parts:

Firstly, we were interested in how many people could use such a P300 based BCI interface at all. A similar study based on motor imagery has proven that around 6 out of 100 naïve subjects participating in a 6 month lasting study setup during a public fair could control a BCI immediately after only 20 minutes of training [13]. Therefore, it was interesting if a similar percentage of the population could use such a P300 control in the single character study.

Secondly, as P300 control allows using rather high degree of freedoms compared to motor imagery based BCIs, a virtual smart home P300 control interface was designed. Here it was of interest if subjects could also control a more complicated interface with high accuracy and speed in the virtual smart home study.

2 Material and Methods

A total of 41 naïve subjects participated in the study. 38 subjects participated in the first study to determine the accuracy of a P300 control by investigating a larger population of subjects. 3 subjects participated in the second experiment for controlling a virtual smart home environment.

The subjects were seated in front of a laptop computer and were instructed not to move and to keep relaxed. Fig. 1B yields the electrode configuration with 8 EEG derivations used for the study. Typical P300 evoked response data for a target from one subject are overlaid at the corresponding electrode positions. The EEG data were acquired with g.USBamp (24 Bit biosignal amplification unit, g.tec medical engineering GmbH, Austria) and 256 Hz sampling frequency. The ground electrode was located on the forehead; the reference was mounted on the right ear lobe. EEG electrodes were made of gold or sintered Ag/AgCl material.

For both experiments the SC speller was selected as for the smart home environment also non quadratic display matrices were developed.

2.1 Single Character Study

Fig. 1A shows the setup for the SC speller. A total of 36 characters and numbers (A, B, … Z; 0, 1, … 9) are displayed in a quadratic matrix on the computer screen. The SC speller highlights each character individually for 60 ms. Between the flashes there is a short dark time of 40 ms where nothing is flashing up. The subject has now the task to look at the character he/she should spell and count how many times the character flashed up. This helps the person to be concentrated on the task. After 15 flashes of each character the signal processing unit calculates the evoked potential and performs a classification to find the character that the subject investigated. Then the flashing sequence starts again and the subject has to look at the next character. The BCI system must be trained firstly on individual EEG data and therefore the subjects were asked to sequentially "write" (or look at) the 5 characters 'W', 'A', 'T', 'E', and 'R'. This process took about 5 minutes.

A B

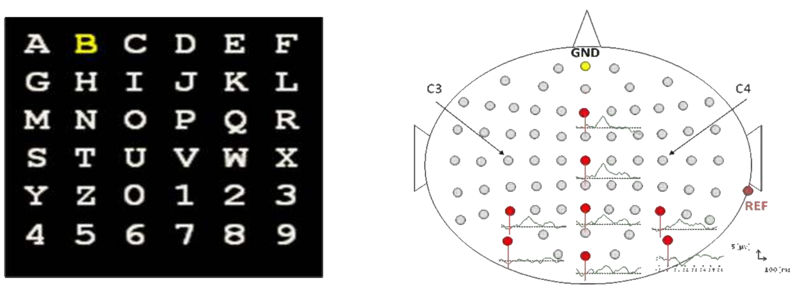


Fig. 1. panel A: Screen layout for the 36 characters; panel B: Top view of head and electrode setup for P300 experiment. The nose is pointing to the top of the page. Electrode (from top left to bottom right indicated as dark shaded disks) positions Fz, Cz, P3, Pz, P4, PO7, Oz and PO8 are used. A typical P300 evoked potential for a target response averaged across 15 trials is indicated near the corresponding electrode positions.

Then the BCI system was trained on the EEG features based on a linear discriminant analyzer (LDA). In the next run the subject had to spell the word 'LUCAS' which took again around 5 minutes.

The Simulink model shown in Fig. 2 is used for the real-time analysis of the EEG data. The *g.USBamp block* reads in the data from the 8 EEG channels. Then data are band pass filtered between 0.5 and 30 Hz in the *Filter block* and down-sampled from 256 Hz to 64 Hz in the *Downsample block* . The S*ignal Processing block* calculates the evoked potentials and performs further averaging across time points and data reduction. A total of 12 feature values per P300 evoked potential is sent to a linear discriminant analyzer for classification. The *Scope* and *To File blocks* are used to visualize the EEG data and to store it for later off-line analysis. The whole Simulink model is driven by the *g.USBamp* hardware block which ensures that the model is updated every 1/256 s. The *Impedance Check block* is utilized to ensure low impedance values for the EEG electrodes.

The *Single Character Speller block* controls the experiment and highlights the corresponding characters randomly. It sends also an identifier *ID-Flash* of the flashing character to the *Signal Processing block*. The *Signal Processing block* generates a buffer for each character and stores the incoming EEG data 100 ms before and 700 ms after the flash occurred (800 ms epoch). This is done until 36 buffers are filled with 15 epochs (15 flashes of each character). Finally an LDA is used to classify the EEG data and to find the buffer closet to the trained P300 response. This classification result yields the character that the subject mentally selected and it will be displayed on the computer screen. Then the next character can be selected by the subject. For offline analysis the time of the *Flash* onsets and the *ID-Flash* of each flashing character are stored.

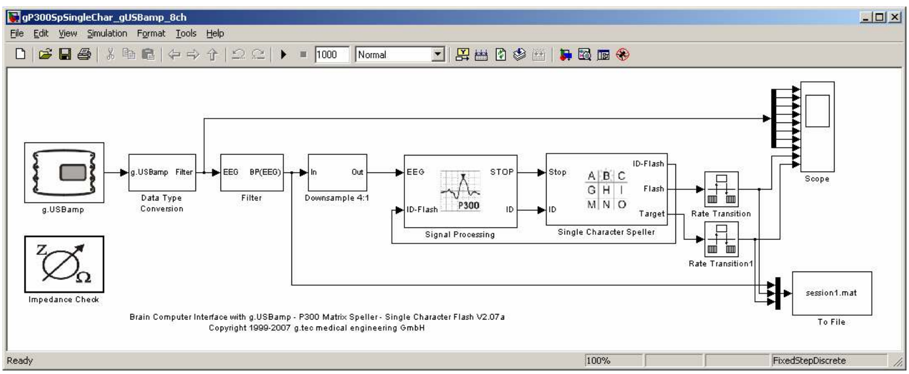


Fig. 2. Simulink model P300SpSingleChar_gUSBamp for the Single Character Speller. See text for the explanation of the different blocks.

2.2 Virtual Smart Home Study

Three subjects participated then in the experiments for smart home control. The electrode setup and recording details were not changed. At the beginning of the experiment the BCI system was trained based on the P300 response of 42 characters of each subject with 15 flashes per character (about 40 minutes training time). All 3 subjects needed between 3 and 10 flashes (mean 5.2) per character to reach an accuracy of 95 % for the single character speller. This resulted in a maximum information transfer rate of 84 bits/s for the single character speller.

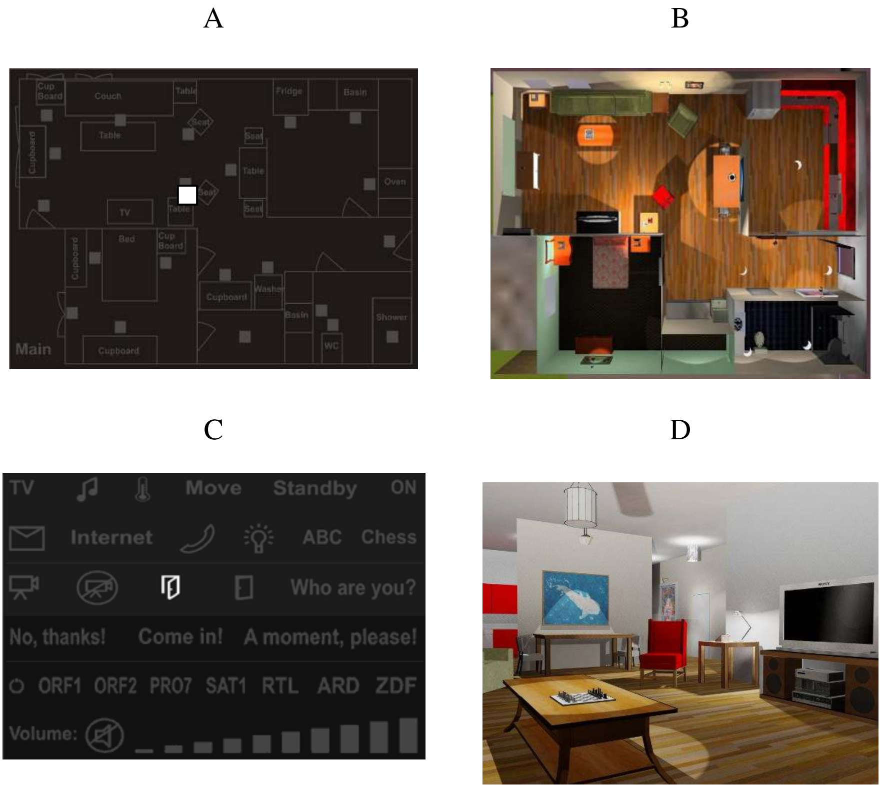


Fig. 3. panel A: Bird view of the apartment; the subject concentrates to the little blinking square at the table to go to the living room; panel B: Bird view of the virtual apartment representation; panel C: Control mask for selecting and controlling features from the TV set; panel D: 3D view of the living room

In the experiment it should be possible for a subject to switch on and off the light, to open and close the doors and windows, to control the TV set, to use the phone, to play music, to operate a video camera at the entrance, to walk around in the house and to move him/herself to a specific place in the smart home. Hence the P300 based BCI system was connected to a Virtual Reality (VR) system. A virtual 3D representation of a smart home with different control elements was developed based on the XVR environment (eXtreme Virtual Reality, University of Pisa). Fig. 3A and 3B yield a bird view of the apartment layout. The upper left panel represents the user interface. The small squares are flashed on and off in a random manner similar to the SC spelling interface. Here the user selected to be set to the living room in a goal orientated way and then selected to operate the TV set. The upper right panel gives the actual representation of the virtual smart home environment. Fig. 3C yields the control masks for controlling the TV set. Fig. 3D yields a 3D view of the living room. The Simulink model in Fig. 4 controls the virtual smart home. The main difference to model *P300SingleCharacterFlash* mode can be found in *Control Flash Smart Home block*.

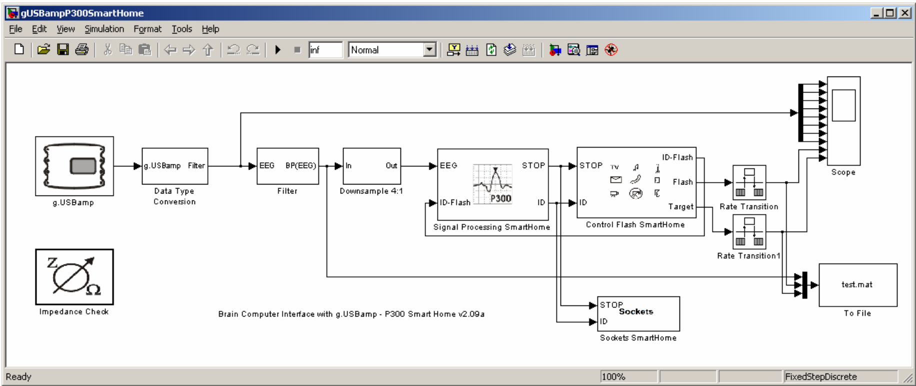


Fig. 4. Simulink model gUSBampP300SmartHome for controlling the virtual environment. See text for the explanation of the different blocks.

The *Control Flash Smarthome block* controls the experiment and highlights the corresponding icons randomly. It sends also an identifier ID-Flash of the flashing icon to the *Signal Processing Smarthome* block. Similar to the single character speller block the signal processing block generates a buffer for each icon and stores the incoming EEG data. Finally an LDA is used to classify the EEG data and to find the buffer closest to the trained P300 response. This classification result yields the icon that the subject mentally selected. This icon represents also the control command for the virtual environment and the command is sent via a the *Sockets SmartHome* block to the XVR smart home representation on a separate PC. Then the next icon can be selected by the subject. For offline analysis the time of the Flash onsets and the Target icon is stored for training of the LDA classifier.

3 Results

3.1 Single Character Study

From all sessions done by the 38 subjects, 55% of the subjects could use the SC speller immediately with 100% accuracy. This means all 5 characters of 'LUCAS' were written correctly. In 76% of the sessions the subjects had only 1 mistake. It must be noted that this is an on-line results and not a cross-validation result. Only in 2.5% of the sessions the subjects were not able to spell a single character correctly. All subjects participated also in a RC study. Results from a RC to SC speller comparison and results from an accompanying questionnaire will be published elsewhere. So the overall accuracy was > 80%.

3.2 Virtual Smart Home Study

Table 2 displays the results of the 3 subjects for the 3 parts of the experiment and for the 7 control masks. Interestingly, the light, the phone and the temperature mask were

Table 1. Percentage of sessions which were classified with certain accuracy. Results based on data from 38 subjects for the single character speller are depicted.

Classification Accuracy in [%]	Percentage of Sessions [N=38]
100	55,3
80-100	76.3
60-79	10.6
40-59	7.9
20-39	2.6
0-19	2.6
average accuracy of all subjects	82

Table 2. Accuracy of the BCI system for each part and control mask of the experiment for all subjects

Mask	Part1	Part2	Part3	Total
Light	100%	100%	100%	100%
Music	-	89,63%	-	89,63%
Phone	-	100%	-	100%
Temperature	100%	-	-	100%
TV	83,3%	-	-	83,3%
Move	88,87%	-	93,3%	91,1%
Go to	100%	-	88,87%	94,43%

controlled by 100 % accuracy. The Go to mask was controlled with 94.4 % accuracy. The worst results were achieved for the TV mask with only 83.3 % accuracy.

Table 3 displays the number of symbols for each mask and he resulting probability that a specific symbol flashes up. If more symbols are displayed on one mask, then the probability of occurrence is lower resulting in increased amplitudes of the P300 responses which should be easier to detect. The flashes column shows the total number of flashes per mask until a decision is made. The translation time per character that is longer if more symbols are on the mask.

Table 3. Number of symbols, occurrence probability per symbol, number of flashes per mask (e.g. 25 x 15 = 375) and conversion time per character for each mask

Mask	Symbols	Propability	Flashes	Time per character [s]
Light	25	4	375	33.75
Music	50	2	750	67.50
Phone	30	3.3	450	40.50
Temperature	38	2.6	570	51.30
TV	40	2.5	600	54.00
Move	13	7.7	195	17.55
Go to	22	4.5	330	29.70

4 Discussion

4.1 Single Character Study

This study showed that the P300 spelling device works with a very high accuracy after only 5 minutes of training. 72.8 % of the subjects were able to spell immediately with 100 % accuracy with the RC speller. This can be compared to an earlier study performed with 99 subjects and motor imagery in Graz [13]. The subjects had to imagine left and right hand movement (20 times each) to move a cursor on the screen to the corresponding side. Then the BCI system was trained on this EEG data (recorded from positions C3 and C4). The training time was also around 6 minutes and the recursive least square or band power estimation in predefined frequency bands with LDA were used for classification. 6.2 % were able to reach accuracy between 90 - 100 % as shown in Table 2. This is well below the P300 results achieved in this study. Of course the motor imagery BCI worked only with 2 bipolar derivations compared to 8 EEG electrodes for the P300 experiment, but the assembly time is almost equal.

4.2 Virtual Smart Home Study

The P300 based BCI system was successfully used to control a smart home environment with accuracy between 83 and 100 % depending on the mask type. The difference in accuracy can be explained by the arrangement of the icons.

However, the experiment yielded 3 important new facts: (i) instead of displaying characters and numbers to the subject also different icons can be used, (ii) the BCI system must not be trained on each individual character, (iii) from all experiments a grand average classifier was built and tested in selected subjects. In contrast to motor imagery BCIs were the system must be retrained every time it is used, the P300 approach can use a standard classifier. The BCI system was trained with EEG data of the spelling experiment and the subject specific information was used also for the smart home control. This allows using icons for many different tasks without prior time consuming and boring training of the subject on each individual icon. This reduces the training time in contrast to other BCI implementations were hours or even weeks of training are needed [1, 2, 3]. This reduction in training time might be important for locked-in and ALS patients who have problems with the concentration over longer time periods. The P300 concept works also better if more items are presented in the control mask as the P300 response is more pronounced if the likelihood that the target character is highlighted drops down [4]. This results of course in a lower information transfer rate, but enables to control almost any device with such a BCI system. Especially applications which require reliable decisions are highly supported. Therefore the P300 based BCI system enables an optimal way for the smart home control. The virtual smart home acts in such experiments as a testing installation for real smart homes.

Also wheelchair control, which many authors identify as their target application, can be realized with this type of BCI system in a goal oriented way. In a goal oriented BCI approach it is then not necessary e.g. to move a robotic hand by thinking about hand or foot movements and controlling right, left, up, down commands. In a more natural way humans just think "I want to grasp the glass" and the real command is initiated by this type of BCI implementation.

Acknowledgments

This work was supported by the EC projects Presenccia and SM4ALL.

References

1. Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kubler, A., Perelmouter, J., Taub, E., Flor, H.: A spelling device for the paralysed. Nature 398, 297–298 (1999)
2. Wolpaw, J.R.: Brain-computer interfaces as new brain output pathways. J. Physiol. 579, 613–619 (2007)
3. Birbaumer, N., Cohen, L.G.: Brain-computer interfaces: communication and restoration of movement in paralysis. J. Physiol. 579, 621–636 (2007)
4. Bayliss, J.D., Ballard, D.H.: A virtual reality testbed for brain-computer interface research. IEEE Transactions on Rehabilitation Engineering 8(2), 188–190 (2000)
5. Pfurtscheller, G., Leeb, R., Keinrath, C., Friedman, D., Neuper, C., Guger, C., Slater, M.: Walking from thought. Brain Res. 1071, 145–152 (2006)
6. Friman, O., Volosyak, I., Graser, A.: Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces. IEEE Trans. Biomed. Eng. 54, 742–750 (2007)
7. Zhang, H., Guan, C., Wang, C.: Asynchronous P300-based brain-computer interfaces: a computational approach with statistical models. IEEE Trans. Biomed. Eng. 55, 1754–1763 (2008)
8. Lin, Z., Zhang, C., Wu, W., Gao, X.: Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs. IEEE Trans. Biomed. Eng. 54, 1172–1176 (2007)
9. Neuper, C., Scherer, R., Wriessnegger, S., Pfurtscheller, G.: Motor imagery and action observation: modulation of sensorimotor brain rhythms during mental control of a brain-computer interface. Clin. Neurophysiol. 120, 239–247 (2009)
10. Donchin, E., Spencer, K.M., Wijesinghe, R.: The mental prosthesis: assessing the speed of a P300-based brain-computer interface. IEEE Trans. Rehabil. Eng. 8, 174–179 (2000)
11. Sellers, E.W., Krusienski, D.J., McFarland, D.J., Vaughan, T.M., Wolpaw, J.R.: A P300 event-related potential brain-computer interface (BCI): the effects of matrix size and inter stimulus interval on performance. Biol. Psychol. 73, 242–252 (2006)
12. Komatsu, T., Hata, N., Nakajima, Y., Kansaku, K.: A non-training EEG-based BMI system for environmental control. Neurosci. Res. 61(Suppl.1), S251 (2008)
13. Guger, C., Edlinger, G., Harkam, W., Niedermayer, I., Pfurtscheller, G.: How many people are able to operate an EEG-based brain-computer interface (BCI)? IEEE Trans. Neural Syst. Rehabil. Eng. 11, 145–147 (2003)

Wearable and Wireless Brain-Computer Interface and Its Applications

Chin-Teng Lin[1,2], Li-Wei Ko[1,2], Che-Jui Chang[1,2], Yu-Te Wang[1,3], Chia-Hsin Chung[1,3], Fu-Shu Yang[1,3], Jeng-Ren Duann[1,4], Tzyy-Ping Jung[1,4], and Jin-Chern Chiou[1,2]

[1] Brain Research Center, National Chiao-Tung University (NCTU), Hsinchu 300, Taiwan, R.O.C.
[2] Department of Electrical and Control Engineering, National Chiao-Tung University, Hsinchu 300, Taiwan, R.O.C.
[3] Department of Computer Science, National Chiao-Tung University, Hsinchu 300, Taiwan, R.O.C.
[4] Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California San Diego, La Jolla, CA, 92093-0961 USA
{ctlin,lwko}@mail.nctu.edu.tw, datocc@hotmail.com, happyhodala@gmail.com, verachung.cs96g@nctu.edu.tw, yfs0606@hotmail.com, duann@sccn.ucsd.edu, jung@sccn.ucsd.edu, chiou@mail.nctu.edu.tw

Abstract. This study extends our previous work on mobile & wireless EEG acquisition to a truly wearable and wireless human-machine interface, NCTU Brain-Computer-Interface-headband (BCI-headband), featuring: (1) dry Micro-Electro-Mechanical System (MEMS) EEG electrodes with 400 ganged contacts for acquiring signals from non-hairy sites without use of gel or skin preparation; (2) a miniature data acquisition circuitry; (3) wireless telemetry; and (4) online signal processing on a commercially available cell phone or a lightweight, wearable digital signal processing module. The applicability of the NCTU BCI-headband to EEG monitoring in real-world environments was demonstrated in a sample study: cognitive-state monitoring and management of participants performing normal tasks.

Keywords: Dry electrodes, brain computer interface, mobile and wireless EEG.

1 Introduction

Electroencephalogram (EEG) is a powerful non-invasive tool widely used by for both medical diagnosis and neurobiological research as it provides high temporal resolution in milliseconds. Another important advantage of EEG is that it involves sensors light enough to allow near-complete freedom of movement of the head and body, making EEG the clear choice for brain imaging of humans performing normal tasks in real-world environments [1]. However, the lack of portable and user-acceptable (e.g., comfortably wearable) sensors and miniaturized supporting hardware/software to continuously acquire and process EEG has long thwarted the applications of EEG

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 741–748, 2009.

monitoring in the workplace [2]. Recently, we developed and tested a prototype four-channel mobile and wireless EEG system incorporating a miniature data acquisition (DAQ) circuitry and dry Micro-Electro-Mechanical System (MEMS) electrodes with 400 ganged contacts for acquiring signals from non-hairy sites without use of gel or skin preparation [2-4]. This study extends our previous work, NCTU BCI-cap [2-4], to a smaller, lighter, wearable & wireless brain-computer interface (BCI), **NCTU BCI-headband**. The NCTU BCI-headband features: (1) disposable dry MEMS electrodes; (2) an 8-channel DAQ unit; (3) wireless telemetry and (4) real-time digital signal processing (DSP) implemented on a commercially available cell phone or a digital signal processing module. The applicability of the NCTU BCI-headband to EEG monitoring in operational environments was demonstrated by a sample study: cognitive-state monitoring and management of participants performing normal tasks in real-world environments.

2 Wearable and Wireless Brain-Computer Interface

Figure 1 shows the system diagram of the mobile and wireless brain-computer interface. The front-end unit integrates (1) clip-on electrode holders for dry MEMS or commercially available wet EEG electrodes, (2) a DAQ unit, and (3) wireless-transmission circuitry, into a quickly and easily donned and doffed **headband** that can acquire and transmit EEG signals from up to eight channels. The back-end unit integrates a wireless signal receiver and on-line DSP. EEG signals are first acquired by dry MEMS or commercially available electrodes, amplified by the preamplifier, converted to digital signals, and then wirelessly transmitted to the data receiver. The DSP unit processes the EEG data and displays the results. The raw EEG data can also be wirelessly transmitted to a remote PC for further offline analysis and/or database collection.

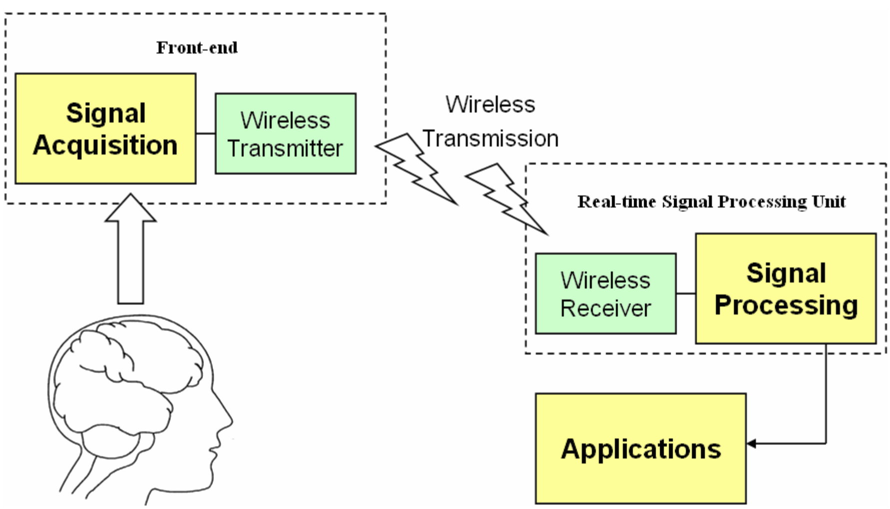


Fig. 1. System diagram of a wearable and wireless brain-computer interface

2.1 Dry MEMS Electrodes and Electrode Holders

We previously explored the use of MEMS technology to build a silicon-based spiked electrode array or so-called dry electrode, to enable EEG, EOG, ECG, and EMG monitoring without conductive paste or scalp preparation [2-4]. However, the connectors between the dry sensors and DAQ board were not very robust in the BCI-cap. This study incorporated snap-on electrode holders to house dry electrodes or commercially available EEG sensors. Fig. 2B shows the snap-on connector.

2.2 Data Acquisition Unit

The data acquisition unit integrated an analog preamplifier, a filter, and an analog-to-digital converter (ADC) into a small, lightweight, battery-powered DAQ. EEG signals are sampled at 512Hz with 12-bit precision, amplified by 6000 times, and band-pass filtered between 1 and 50 Hz. Fig. 2A shows the block diagram of the DAQ unit. Fig. 2B shows the DAQ unit for each electrode (20mm x 18mm PCB 'node'). To reduce the number of wires for high-density recordings, the power, clocks, and measured signals are daisy-chained from one node to another with bit-serial output. That is, adjacent nodes (electrodes) are connected together to (1) share the power, reference voltage, and ADC clocks, and (2) daisy chain the digital outputs.

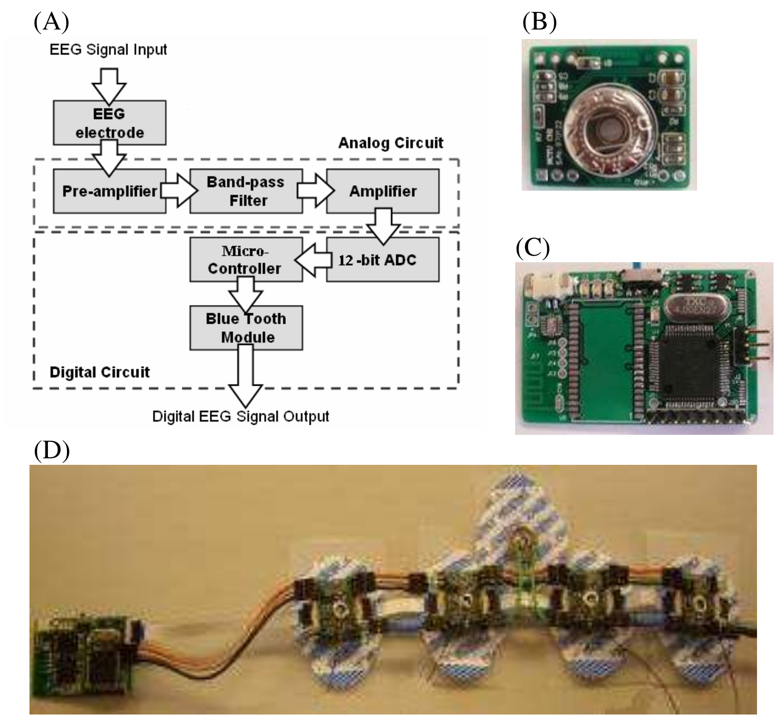


Fig. 2. (A) Block diagram of the data acquisition unit, (B) the DAQ unit for each electrode, (C) the wireless transmission unit, and (D) the integrated circuits of the NCTU BCI-headband

2.3 Wireless Transmission Unit

The wireless-transmission unit consisted of a wireless module and a micro-controller. It used a Bluetooth module to send the acquired EEG signals to a custom real-time DSP unit described below or a Bluetooth-enable cell phone which was used as a real-time signal-processing unit. The dimension of the wireless transmission circuit was 40 x 25 mm^2 (as shown in Figure 2C). Figure 2D shows a picture of the integrated 4-channel wireless EEG system. A reference and a ground channels were also included in the system (not shown). The integrated circuitry can be embedded into a headband, NCTU BCI-headband, as shown in Figure 3. The power-consumption of the NCTU BCI-headband is very low (a 1100 mAh Li-ion battery can last over 33 hours).

Fig. 3. A picture of the wearable & wireless EEG system, NCTU BCI-headband. It comprises 4- or 8-channel snap-on electrode holders (plus a reference and a ground channels), miniature bio-amplifier, a bandpass filter, an ADC and a Bluetooth module. All channels were referred to the left mastoid.

2.4 Real-Time Digital Signal Processing Unit

To be practical used in operational environments, the signal processing unit must be light-weight, portable, low-power, and have on-line data receiving and real-time signal processing function. Therefore, this study designed and developed a real-time digital signal processing unit which used a Bluetooth module to receive the acquired EEG signals from the NCTU BCI-headband and process the EEG signals via its core processor in near real-time. The core processor is the Blackfin processor (Analog Device Incorporation, ADSP-BF533) which provided a high performance, power-efficient processor choice for demanding signal processing applications. The dimension of the miniature DSP unit is about 65 x 45 mm^2 (as shown in Figure 4).

The maximum high processing performance of the BF533 core processor can reach 600MHz. Furthermore, the following peripheral modules were also incorporated in the unit.

- SD RAM and FLASH memory
- RS-232 serial interface
- Six keypads and a LCD panel (240 by 320 pixels)
- JTAG interface for debug and FLASH programming

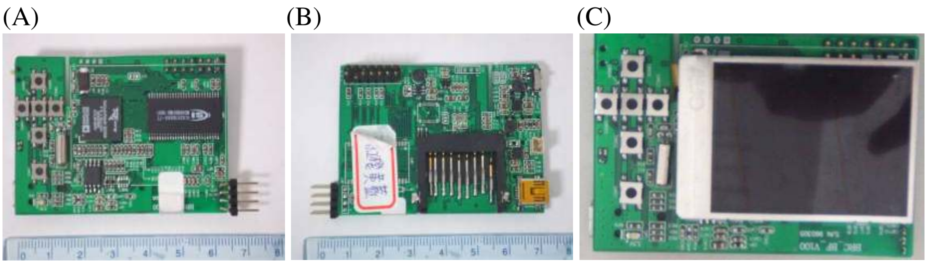


Fig. 4. Real-time digital signal processing unit. (A) Front panel houses an ADI BF533 processor and six keypads, (B) Back panel houses a SD card adapter, a Bluetooth module and a USB module, (C) A LCD is mounted on the frontal panel of the DSP unit to display the received raw data or the results of DSP.

- Bluetooth module
- USB chargeable and programming module

2.5 Data-Logging and Digital Signal Processing on a Cell Phone

To demonstrate the application of the wearable & wireless BCI during long and routine recording in operational environments, we have developed and installed a data-logging Java graphical user interface (GUI) on a Bluetooth-enable cell phone. The Java program receives EEG signals from the NCTU BCI-headband and plots them on the LCD screen. We have also implemented power spectrum density (PSD) estimation using a 512-point Fast Fourier Transform (FFT) on the cell phone.

3 Testing of Nctu Bci-Headband

3.1 Comparison between the NCTU BCI-Cap and BCI-Headband

Lin et al. [2] reported a 4-channel BCI-cap which measured EEG and transmitted it to a commercially available DSP kit by Texas Instruments. This study extended the EEG recording system into a truly mobile brain-computer interface which acquired and processed EEG signals in near real-time. Table 1 compares the specifications and features of the BCI-cap and those of the BCI-headband. It is evident that, compared to BCI-cap, BCI-headband is lighter, smaller, more power-efficient and accommodates more channels with higher sampling rate and digitization precision.

3.2 Real-Time Alertness Monitoring Using NCTU BCI-Headband and a Cell Phone

Lin et al. [2, 5] recently demonstrated the feasibility of using dry MEMS EEG electrodes, supporting hardware and commercially available TI DSP kit to continuously and accurately estimate the driving performance (putative drowsiness level) based on EEG data from four frontal non-hairy positions in a realistic VR-based dynamic driving simulator. This study implemented the cognitive-state monitoring algorithm on a

Table 1. Comparison between the NCTU BCI-cap [2] and BCI-headband

	NCTU BCI-cap [2] 	**NCTU BCI-headband** 
Dimension (mm)	 46 x 66 mm^2	 DAQ: 20 x 18 mm^2 (4 pieces)  Wireless Unit: 40 x 25 mm^2
Weight	185 g	< 100 g
Precision	8 bit	12 bit
Sampling Rate	200Hz	512Hz
Bandpass Filter	1 - 50 Hz	
Gain	5000 times	6000 times
Output Current	480 mA	31.58 mA
Battery Life (3.7V, 1100mAh)	**3-4 hours**	**33-34 hours**

Bluetooth-enable cell phone that received EEG signals and processed them with the on-board processor. The cell phone delivered arousing feedback when the participants were drowsy. Figure 5A shows the flowchart of the signal processing implemented on the cell phone. Figure 5B shows the evident alpha activities when the subject was drowsy. We have also developed a user-friendly GUI (the pie chart in Figure 5C) to

(A) (B)

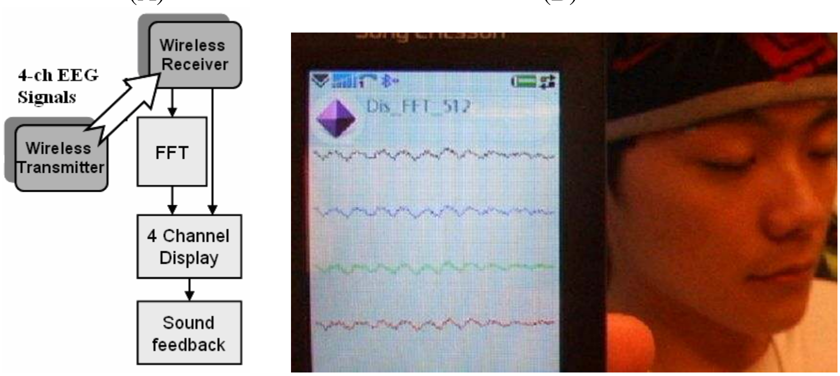


Fig. 5. (A) The flowchart of signal processing on a cell phone, (B) Four-channel EEG signals were displayed on the cell phone. Alpha rhythm became evident when the subject was drowsy. (C) A custom GUI of BCI-based cognitive-state monitoring on a cell phone.

(C)

Fig. 5. (*continued*)

continuously track and display the cognitive states of the subject. When the subject was alert, the whole pie chart was blue. As the subjects became more and more drowsy, wedges of the pie chart changed from blue to red, increasing the red area as the subject became drowsier. When the subject became very drowsy, the whole pie chart became red and the warning set off.

4 Conclusions

This study demonstrated a truly portable, lightweight, and readily wearable brain-computer interface that featured dry MEMS electrodes and a miniaturized DAQ, wireless telemetry and online signal processing. The main goal of the design and development of wearable and wireless BCI is to maximize their wearability, unconstrained mobility, usability and reliability in operational environments. In this study, the signal-processing module and the Bluetooth-enable cell phone were programmed to assess fluctuations in individuals' alertness and capacity for cognitive performance based on the EEG signals. The BCI delivered arousing feedback to the driver to maintain optimal performance. The cell phone and DSP unit, however, can be programmed for many other brain-system interface applications. We expect that a truly portable and user-acceptable BCI will have enormous future impacts on clinical research and practice in neurology, psychiatry, gerontology, and rehabilitation medicine.

Acknowledgement. This work was supported by the National Science Council, Taiwan, R.O.C., under Contracts NSC 97-2627-E-009-001 and NSC NSC 97-2220-E-009-052; the "Aiming for the Top University Plan," National Chiao-Tung University, the Ministry of Education, Taiwan, under Contract 97W806; and U.S. Army Research Laboratory under Grant W911NF-08-2-0003. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation heron.

References

1. Makeig, S., Gramann, K., Jung, T.-P., Sejnowski, T.J., Poizner, H.: Linking brain, mind and behavior: The promise of mobile brain/body imaging. International Journal of Psychophysiology (in press)
2. Lin, C.T., Ko, L.W., Chiou, J.C., Duann, J.R., Huang, R.S., Chiu, T.W., Liang, S.F., Jung, T.P.: Noninvasive neural prostheses using mobile & wireless EEG. Proceedings of the IEEE 96(7), 1167–1183 (2008)
3. Lin, C.T., Chen, Y.C., Huang, T.Y., Chiu, T.T., Ko, L.W., Liang, S.F., Hsieh, H.Y., Hsu, S.H., Duann, J.R.: Development of wireless brain computer interface with embedded multitask scheduling and its application on real-time driver's drowsiness detection and warning. IEEE Transactions on Biomedical Engineering 55(5), 1582–1591 (2008)
4. Chiou, J.C., Ko, L.W., Lin, C.T., Jung, T.P., Liang, S.F., Jeng, J.L., Hong, C.T.: Using novel MEMS EEG sensors in detecting drowsiness application. In: IEEE Biomedical Circuits and Systems Conference (BioCAS 2006), London, England, November 29-December 1 (2006)
5. Lin, C.T., Wu, R.C., Jung, T.P., Liang, S.F., Huang, T.Y.: Estimating driving performance based on EEG spectrum analysis. EURASIP Journal on Applied Signal Processing 19, 3165–3174 (2005)

Mind Monitoring via Mobile Brain-Body Imaging

Scott Makeig

Swartz Center for Computational Neuroscience, Institute for Neural Computation,
University of California, San Diego, USA
scott@sccn.ucsd.edu

Abstract. Current brain-computer interface (BCI) research attempts to estimate intended operator body or cursor movements from his/her electroencephalographic (EEG) activity alone. More general methods of monitoring operator cognitive state, intentions, motivations, and reactions to events might be based on continuous monitoring of the operator's (EEG) as well as his of her body and eye movements and, to the extent possible, her or his multisensory input. Joint modeling of this data should attempt to identify individualized modes of brain/body activity and/or reactivity that appear in the operator's brain and/or behavior in distinct cognitive contexts, if successful producing, in effect, a new *mobile brain/body imaging* (MoBI) modality. Robust MoBI could allow development of new *brain/body-system interface (BBI)* designs performing multidimensional monitoring of an operator's changing cognitive state including their movement intentions and motivations and ('top-down') apprehension of sensory events.

Keywords: cognitive monitoring; electroencephalography (EEG); motion capture; independent component analysis (ICA); brain-computer interface (BCI), mobile brain/body imaging (MoBI); human-computer interface (HCI).

1 Introduction

Over the last decade there has been an explosion of interest in using EEG to monitor selected movement intentions of an operator trained to produce changes in the amplitude of one or more EEG measures that are mechanically associated by a brain-computer interface (BCI) system with two or more intended external actions (in simplest form, moving a screen cursor up or down). BCI research was first funded to construct systems allowing communication by a relatively few cognitively intact but totally paralyzed or 'locked-in' subjects though, naturally, first exploratory phases of BCI research use normal test subjects. To insure the possibility that the methods developed in these phases might be usable by the target locked-in subjects, it was important to establish that the EEG changes used to detect movement intentions were not based on non-brain contributions to EEG signals recorded on the scalp, e.g., activity arising from subject eye movements or scalp muscle activities. Thus, for many researchers the BCI concept became identified with the goal of using 'pure' EEG, apart from non-brain 'artifacts,' to convey and decipher a subject's stereotyped cursor (or body) movement intentions.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 749–758, 2009.

The goal of providing a useful, non-invasive communication system for 'locked-in' subjects is surely laudable, and actual demonstrations that both a few 'locked-in' and many normal subjects can communicate (albeit quite slowly) via learned control of their macroscopic brain activity patterns, without involvement of direct motor control, are novel and intriguing. However, unnecessary adherence to this limited BCI goal could slow development of more general classes of human-system interfaces involving continuous monitoring of non-invasively recorded brain activity.

1.1 Unexplored Problems in BCI Research

As a new subject, at least four fundamental questions about the operation, limitations, and effects of EEG-based BCI operation remain unexplored:

1. Key obstacles to widespread acceptance and application of non-invasive EEG-based BCI systems are the need for a long training regimen, and the failure of a significant fraction of subjects to achieve stable, non-invasive BCI control even after intensive training. Finding specific reasons for these difficulties, and methods around them, are fundamental if BCI or more general 'neurotechnology' or 'neuro-ergonomic' HCI research is to have broad applications.
2. When a subject in a BCI experiment learns to move a computer screen cursor by increasing or reducing the amplitude of a selected brain rhythm – whether a mu rhythm, near-DC potential, or other phenomenon – what 'body part' (or brain system) do they use to willfully effect the modulation? While this is a fundamental issue for BCI research, it is one that has so far been nearly ignored.
3. Although achieving volitional control of a BCI system through EEG modulation alone is an intriguing goal, more general questions for HCI systems involving EEG monitoring are how to combine EEG analysis with concurrent recording and analysis of subject behavior, eye and muscle activities, and multisensory input to monitor and adapt to changing human cognitive state, intent, and reactivity.
4. Another relatively unexplored question is whether there are psychobiological effects of training and performing volitional control of natural brain rhythms. These effects might either be phasic (affecting the operator only during BCI operation) or tonic (also affecting their behavior and/or brain activity at other times); they might be positive (for example producing a useful strengthening of attentional control), or negative (some unforeseen consequence of disrupting natural, non-conscious modes of dynamic brain regulation).

All these questions should and must eventually be addressed by the advancing fields of human neuroscience and neurotechnology. This paper discusses a general plan of approach to the first three questions above – How can learning of EEG-based volitional control be made quicker and more universal? What EEG modulatory systems do successful BCI subjects use to learn and to effect volitional control of their EEG activity? And, how can EEG be combined with other information about operator behavior and sensation to allow human-system interactions to estimate and use information about operator mental state and cognitive reactions to events?

1.2 EEG Modulation

EEG dynamics have long been characterized by their diverse spectral profiles. For example, slow semi-rhythmic activity is characteristic of EEG in deep sleep, while awake/alert EEG contains more high-frequency activity. Narrow-band brain rhythms appear most predominantly in the (8-12 Hz) alpha band, but also at somewhat higher and lower frequencies. Spectral modulations of EEG activity at lower and higher frequencies affect broader frequency bands. A considerable (if insufficient) amount is known about several brain systems that modulate the spectrum of local field activity in the brain's cortex, the brain source of most scalp-recorded EEG. A number of these systems are the brainstem-centered 'evaluation' systems labeled by the specific neurotransmitter they project quite widely (acetylcholine, dopamine, norepinephrine, serotonin, or etc.). However, evidence for the involvement of these or other systems in successful BCI control has not been presented.

1.3 Mobile Brain/Body Imaging (MoBI)

The fundamental purpose of the brain is to control behavior or more exactly, to optimize the *outcome* of behavior – maximizing its ensuing rewards and/or minimizing ensuing penalties as per subject purposes, needs, and desires. It is now possible to record brain activity at relatively high bandwidth – a Mbit/sec or more of EEG, MEG, BOLD, single-cell spike/field data, etc. Surprisingly, however, there has been little serious effort to concurrently record the *behavior* the brain is controlling with anything near the same bandwidth. In human brain experiments, behavior is most often recorded only in the form of a sparse series of minimal finger button presses – giving an effective rate of behavioral data collection near 1 bit/sec. Simply from this ~1,000,000:1 mismatch, it is no wonder that recent progress in human psychophysiology has been relatively slow.

The obvious remedy for this oversight is to simultaneously record as much behavioral information as possible in paradigms including some range of natural behavior. It should be desirable to record as wide and natural a range of behavior as possible, in as physically free and natural a behavioral environment as possible. Currently, this goal can only be approached only using EEG brain imaging, since its sensors, alone among current high-bandwidth brain imaging modalities, are light enough that its recording does not require major restriction on subject head or body movements.

Recently, I have proposed the combination of wearable, high-density EEG and body motion capture (combined, ideally, with eye gaze and audiovisual scene recording) may constitute a new brain imaging modality, 'Mobile Brain/Body Imaging' or MoBI [1]. Once successfully developed and demonstrated, MoBI could allow, for the first time, the study of macroscopic brain dynamic patterns supporting natural and naturally motivated actions (and interactions) in normal 3-D environments.

A key first problem to be overcome in realizing the promise of mobile brain imaging is the problem of separating the activities of brain EEG sources from non-brain artifacts, particularly head and neck muscle activities and artifacts induced in the EEG by eye movements. A workable solution to this problem, at least, is the introduction of independent component analysis (ICA) of EEG data [2-6]. Under favorable circumstances, ICA cleanly separates brain and non-brain data source activities that are mixed by volume conduction in scalp EEG recordings, a process for which much open-source software is now available [7]. A second problem is to model the muscular forces producing the observed motor behavior; for this, open-source biomechanical modeling software is also becoming available [8]. Finally, adequate statistical signal processing or machine learning methods are required to discover dynamic links between concurrent brain source activities, muscle activations, and other classes of MoBI data.

Supposing the near-future availability of viable MoBI recording and analysis methods, we can ask how the concept of BCI can be expanded to consider *brain/body interface (BBI)* designs that acquire and continuously update information about the cognitive state, reactions, intentions, and motivations of the system operator from joint MoBI recording.

2 Brain/Body Interface (BBI) Methods

For a BBI system to be maximally effective, it would seem wise to consider and test two design principles:

a) To best understand the complex associations of ongoing multidimensional changes in EEG dynamics with cognitive state, perceptual events, and movement intentions and motivations, the analysis should both observe and take into account the subject's movements (including limb, body, and eye movements), and any other available physiological signals. In other words, to optimally model brain activity it is important to take in to account, as much as possible, the behavior the brain is controlling. This suggests the potential importance of the development of concurrent brain/body imaging recording and analysis, as in the MoBI concept.
b) The information about cognitive state and action motivations and intentions that may be most robustly decoded from joint EEG and behavioral information should concern distinctions between circumstances and events in which EEG dynamics exhibit spontaneous differences. In particular, it is likely that learned control of EEG signals will be most successful when the learned repertoire of EEG modulations used to decode subject control intentions are identical or close to the subject's repertoire of spontaneous EEG modulations.

The identified EEG dynamics used in BBI monitoring and control may either index brain dynamics that play supporting roles in these circumstances, or their cortical source activities may also play a direct role in shaping the joint timing of distributed neural activities, a concept that is gradually being re-introduced into neuroscience by new evidence and by theoretical considerations of the utility of mass action in the central nervous system for controlling behavior and its outcomes.

2.1 EEG Modulators

Standard methods for analyzing EEG data are based on averaging measures of EEG dynamics across trials or time windows, thereby collapsing the continuously time-varying signals into a average representation of activity time-locked to one or more *classes* of events. Further, most EEG analyses focuses on the individual scalp channel signals, though they are each differently weighted mixtures of many brain and non-brain source signals. Independent component analysis attempts to locate discrete sources of information in multidimensional data in which several independent information streams are linearly mixed in sensor data. However, the spectrum of each identified brain source component signal, like every recorded scalp signal, varies irregularly over time. Standard methods for analyzing either independent component or scalp channel signals during a period of continued subject task performance typically model the exhibited variability as noisy deviations from a stable mean spectrum or stable event-related spectral perturbation (ERSP) time/frequency mean, variation noise in which spectral power at each frequency is implicitly assumed to vary independently.

An alternate approach assumes that the observed power spectral variability sums variations in several to many modes of spectral variability (and co-variability) that are characteristic of the component source process. Earlier, we introduced the device of converting component spectrograms to log power while positing that the motive force behind these modal modulations are processes that modulate spectral activity *multiplicatively,* at characteristic frequencies, with independent or near-independent time courses or effect distributions across trials [9]. Recently, we have tested the use of ICA decomposition the ongoing log power spectrograms of a number of independent component processes from single subjects performing eyes-closed imagination exercises[1]. Log spectral decomposition separated second-to-second variations in the log spectrogram into a log sum of multiplicative modulator processes, each with a fixed spectral and spatial component effect template whose effect on the affected spatial component log spectra is determined by multiplication by a single log amplitude time series. This approach gave a number of interesting results including alpha band processes at different frequencies plus harmonics, broader beta and theta band processes, and very broadband shifts in power distribution.

We have also experimented with adding information to the analysis about the time locking and other experimental events and the context in which they occur. The goal of this analysis approach is to avoid so far as possible *the method of planned comparisons*, the basis for most experimental data analysis, in which measures for pairs of conditions are compared, each measure an identically weighted average of measurements characterized by one (or sometimes more) key variable value.

For example, there have been thousands of EEG studies that compared the average responses (typically called ‘P300’) to ‘target’ and ‘non-target’ stimuli in a simple attention task. The underlying assumption here is that the brain emits identical responses to each ‘target’ or ‘non-target’ stimulus, respectively, regardless of the local event context. The hope is that the effect of the ‘target/non-target’ variable is separable from other variables, and essentially stable across time. Unfortunately, this is not the case. P300 ‘target’ responses vary widely in amplitude and scalp distribution from trial to trial, and this type of variability limits the performance of simple BCI systems, for example one that might attempt to set a fixed threshold to identify the appearance of a ‘target’ response, regardless of event context [10].

[1] Onton, J. and Makeig, S., unpublished data.

I propose that BBI research explore an alternative approach in which multiple characteristic relationships between EEG dynamics and single *events in context* are determined directly *from* the joint EEG, stimulus, and behavioral data. Some facts concerning the nature of individual events may be available to at BBI system in real time, for example the moment and screen on which a piece of information is presented, or the screen to which the subject is directing their gaze.

An example of an *unavailable* context variable might be the interpretation of the subject of a visual event as representing a challenge or threat. In pilot data recorded to build an individualized (or collective) BBI model, the level of threat could be varied systematically and the level of perceived threat might be estimated from the subject's brain and behavioral responses. In subsequent real-time operation, other variables defining the current event and event context may be available from the system event log and subject behavioral record.

Combined with direct observation of the EEG and subject motor behavior, these *available* context variables, combined, may allow estimation of the *unavailable* variable – here, whether and to what extent the subject perceives a visual event to signal a threat to the operation of the system. This information might be used to immediately deploy available additional countermeasures whenever a genuine perceived system threat is estimated to occur, or possibly to monitor the state of responsiveness of the subject when false indications of (test) threats are delivered to the subject, probing their advancing level of expertise in recognizing a threatening event, or for estimating their current cognitive fitness for duty.

If the system response to the operator's appraisal of a threatening event helps the operator mount an adequate and timely response, then the system response will serve as a powerful reward, and naturally over time and use the operator's EEG pattern should be expected to adapt to give a more distinct perceived-threat signal to the system. Thus, a natural cognitive response monitoring system could easily become an interactive learned BCI/BBI system. Further, it is natural to hypothesize that when the system is based on the operator's natural brain response modes, it may also be natural and relatively easy for the operator to learn to produce the EEG patterns that are most distinctly and reliably detected by the system.

Fig. 1 gives the gist of the concept in graphic form. Three types of MoBI data may be recorded concurrently to run a brain/body interface (BBI): high-density EEG data, behavioral data, and context data (event information, event, EEG, and behavioral history, etc.). Standard BCI systems (*dotted arrows*) attempt to estimate some parameter of the behavioral and/or event/context data directly from the scalp EEG using a machine learning approach. In the proposed BBI model (*wide light blue arrows),* the EEG data are first separated into cortical (and non-brain) EEG source processes (*thin blue arrows*), the spectral modulator processes operating on these source processes are estimated from the EEG source data, and the linkage of the EEG source and source modulator processes to the behavioral and/or event/context data are determined. When one or more parameters of the event/context data are unavailable (e.g., in real-time operation), any of the available MoBI data may be used in a BBI to estimate the unavailable parameter. The estimation process might be designed to perform well even when additional data variables are missing. The MoBI data used for this estimation might include available behavioral data (body motion capture, eye gaze tracking, etc.) and event/context information as well as EEG dynamics.

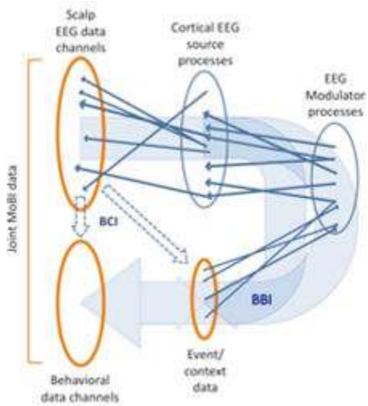


Fig. 1. Schematic model diagram for a non-invasive brain/behavior system interface (BBI) design. Concurrent scalp EEG, behavior, and event/context data are collected in a Mobile Brain/body Imaging (MoBI) paradigm (*thick ovals*). In most currently proposed BCI systems (*dotted arrows*), selected EEG data are processed in near-real time to estimate or predict some behavioral or event/context parameter (‘*BCI*’). In the proposed BBI, the EEG data are first separated into cortical EEG source processes (*upper middle oval*) (plus non-brain artifact processes, not shown). Then the time/frequency behaviors of the source processes are further separated into effects of a number of maximally distinct EEG source modulator processes (*upper right oval*). In the BBI model, both selected EEG time-domain source and frequency-domain source modulator data may be integrated with the behavioral and other event/context data to estimate or predict selected behavioral and/or event/context parameters (*broad light blue arrows*).

3 Discussion

The model of an EEG-based BBI system shown in Fig. 1 has the advantage of involving volitional control of spatiotemporal EEG dynamic patterns most specifically associated with the operator’s spontaneous EEG responses in the targeted event categories [11]. While it is natural to hypothesize that strengthening and controlling spontaneously active EEG patterns may be more easily and quickly learned, this assumption may prove incorrect in some or many cases, and thus basic experiments (and adequate analyses) are needed to test it. Earlier, we showed that applying even highly overlearned BCI control of a single pre-defined EEG feature may involve complex and asymmetric EEG changes in and among many cortical regions [12]. Thus, gaining a basic understanding of the nature and learning of volitional EEG control may in many cases prove to be a complex and difficult process.

How may we determine which brain modulatory systems are involved in spontaneous and learned control of particular EEG or behavioral/EEG dynamics? A full answer to this question may require invasive experiments (potentially involving patient volunteers who have been implanted with cortical electrodes for clinical purposes), positron emission tomography (PET) experiments that can assess neurotransmitter distributions in the brain, various psychopharmacological manipulations, combined with carefully selected behavioral paradigms, for example those directly manipulating reward levels known to be linked to dopamine release [13]. However, a number of brain modulatory systems may be involved in most state changes and event responses of interest, so this investigation should be expected to be involved.

A possible objection to the model shown in Fig. 1 is that if an adequate BCI function linking the recorded EEG signal to the target behavioral or event/context parameter(s) of interest proves to be *linear*, then constructing a more elaborate BBI function linking EEG data first to EEG sources, then to their natural modes of spectral modulation, and finally to the estimated event/context or behavioral measure may not give a better-performing estimator. The proposed EEG source modulator model, however, is nonlinear as it operates on source (log) power spectra. Linear or other functions of the estimated source and source modulator time courses, therefore, involve additional information and might well have advantages over direct (and particularly, linear) BCI estimation. However, use of power spectral estimates ignore source signal phase and with it, precise latency information available in the time-domain data. Thus, applying a joint linear (or other) estimator to combine time-domain and time/frequency-domain data could improve performance over a time-domain estimator alone.

Recently Bigdely Shamlo and colleagues demonstrated a successful application of such an approach [14]. We reported a method for estimating the probability that a briefly presented visual image contained a rare target feature – an airplane feature in a stream of satellite ground images presented to the subject at a rate of 12 images per second. Near-real time performance in correctly detecting the presentation of single target-bearing images solely from high-density EEG (by combining source time-domain and source spectral modulator information in a linear estimator) was high, giving an area under the ROC curve of over 90% for most subjects.

Like most BCI projects, this project did not expressly capture subject behavioral information. However, it did allow use of maximally independent EEG sources capturing potentials induced by characteristic subject eye gaze behavior following target appraisal, unlike BCI systems built to serve completely paralyzed subjects. Although the very rapid serial visual presentation (RSVP) did not reward normal saccadic eye movements, independent components accounting for eye movements following target perception was found to carry some target classification information (though of less value compared to several brain EEG source responses).

The BBI model shown schematically in Fig. 1 does not propose a method for combining EEG and behavioral data, in particular body motion capture data. This is a topic that both requires and deserves much attention and exploration. Of particular interest is to determine the extent to which it is desirable to solve the biomechanical inverse problem, estimating which muscle actions produce the observed sequence of body movements, before estimating links between EEG source activities, body movements, and operator mental state or reactions [8].

Finally, can the proposed MoBI-based BBI systems be practical for widespread application, or must they remain basic research tools? EEG spatial filtering requires the availability of a relatively high number of scalp EEG recording channels. Typically, BCI designers have attempted to maximize signal to noise ratio by restricting the number of channels used in the classifier, an approach that might also lower the cost of the system, if realized using currently available technology. To date, body motion capture (mocap) systems also remain quite expensive. Thus, can the proposed MoBI-based BBI systems become practical for routine application, even in (e.g.)

high-value military or civilian environments? Here, the rapid progress of electronic fabrication methods, allow microminiaturized data acquisition and processing units based on flexible thin-film technologies should allow development and relatively low-cost deployment of wearable high-density EEG and behavioral monitoring systems within a few years [15]. Such systems should be readily applicable to some important problems, for example alertness monitoring of shift-work operators of high-value, high-risk systems [16]. Full realization of the MoBI-based BBI concepts discussed here will likely require a great deal more basic and applied research in many laboratories combining expertise in several fields of neuroscience, mathematics, and engineering.

References

1. Makeig, S., Gramann, K., Jung, T.-P., Sejnowski, T.J., Poizner, H.: Linking brain, mind and behavior: The promise of mobile brain/body imaging (MoBI). Int. J. Psychophysiolog. (accepted for publication)
2. Makeig, S., Jung, T.-P., Ghahremani, D., Bell, A.J., Sejnowski, T.J.: Blind separation of auditory event-related brain responses into independent components. Proc. Natl. Acad. Sci. USA. 94, 10979–10984 (1997)
3. Makeig, S., Westerfield, M., Jung, T.-P., Covington, J., Townsend, J., Sejnowski, T.J., Courchesne, E.: Functionally independent components of the late positive event-related potential during visual spatial attention. J. Neurosci. 19, 2665–2680 (1999)
4. Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., McKeown, M.J., Iragui, V., Sejnowski, T.J.: Removing electroencephalographic artifacts by blind source separation. Psychophysiolog. 37, 163–178 (2000)
5. Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., Sejnowski, T.J.: Analysis and visualization of single-trial event-related potentials. Human Brain Mapping 14, 166–185 (2001)
6. Makeig, S., Westerfield, M., Jung, T.P., Townsend, J., Courchesne, E., Sejnowski, T.J.: Dynamic brain sources of visual evoked responses. Science 295, 690–694 (2002)
7. Delorme, A., Makeig, S.: EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Meth. 134, 9–21 (2004)
8. Delp, S.L., Anderson, F.C., Arnold, A.S., Loan, P., Habib, A., John, C.T., Guendelman, E., Thelen, D.G.: OpenSim: Open-source software to create and analyze dynamic simulations of movement. IEEE Trans. Biomed. Eng. 54, 1940–1950 (2007)
9. Makeig, S., Delorme, A., Westerfield, M., Jung, T.-P., Townsend, J., Courchesne, E., Sejnowski, T.J.: Electroencephalographic brain dynamics following manually responded visual targets. PLOS Biol. 2, 747–762 (2004)
10. Onton, J., Delorme, A., Makeig, S.: Frontal midline EEG dynamics during working memory. NeuroImage 27, 341–356 (2005)
11. Makeig, S., Enghoff, S., Jung, T.-P., Sejnowski, T.J.: A natural basis for efficient brain-actuated control. IEEE Trans. Rehab. Eng. 8, 208–211 (2000)
12. Delorme, A., Makeig, S.: EEG changes accompanying learned regulation of 12-Hz EEG activity. IEEE Trans. Neural Sys. Rehab. Eng. 2, 133–136 (2003)
13. Arias-Carrión, O., Pöppel, E.: Dopamine, learning and reward-seeking behavior. Act. Neurobiol. Exp. 67, 481–488 (2007)

14. Bigdely-Shamlo, N., Vankov, A., Ramirez, R.R., Makeig, S.: Brain activity-based image classification from rapid serial visual presentation. IEEE Trans. Neural Syst. Rehabil. Eng. 16, 432–441 (2008)
15. Chin-Teng Lin, C.-T., Ko, L.-W., Chiou, J.-C., Duann, J.-R., Huang, R.-S., Liang, S.-F., Chiu, T.-W., Jung, T.-P.: Noninvasive neural prostheses using mobile and wireless EEG. Proc. IEEE 96, 1167–1183 (2008)
16. Makeig, S., Inlow, M.: Lapses in alertness: Coherence of fluctuations in performance and the EEG spectrum. Electroencephalogr. Clin. Neurophysiol. 86, 23–35 (1993)

Utilizing Secondary Input from Passive Brain-Computer Interfaces for Enhancing Human-Machine Interaction

T.O. Zander[1,2], C. Kothe[1], S. Welke[1,2], and M. Roetting[1,2]

[1] Department of Psychology and Ergonomics, Chair for Human-Machine Systems, Team PhyPA, Berlin Institute of Technology, Berlin, Germany
[2] Center of Human-Machine Systems, GRK prometei, Berlin Institute of Technology, Berlin, Germany
thorsten.zander@mms.tu-berlin.de

Abstract. A Brain-Computer Interface (BCI) directly translates patterns of brain activity to input for controlling a machine. The introduction of methods from statistical machine learning [1] to the field of brain-computer interfacing (BCI) had a deep impact on classification accuracy. It also minimized the effort needed to build up the skill of controlling a BCI system [2]. This enabled other fields of research to adapt methods from BCI research for their own purposes [3, 4]. Team PhyPA, the research group for physiological parameters of the chair for Human-Machine Systems (HMS) of the Technical University of Berlin, focuses on enabling new communication channels for HMS. Especially the use of passive BCIs (pBCI) [3, 4], not dependent on any intended action of the user, show a high potential for enhancing the interaction in HMS [5]. Additionally, as actual classification rates are still below the threshold for efficient primary control [6, 7] in HMS, we focus on establishing a secondary, BCI-based communication channel. This kind of interaction does not necessarily disturb the primary mode of interaction, providing a low usage cost and hence an efficient way of enhancement. We have designed several applications following this approach. Here we are going to present briefly the results from two studies, which show the capabilities arising from the use of passive and secondary BCI interaction. First, we show that a pBCI can be utilized to gain valuable information about HMSs, which are hard to detect by exogeneous factors. By mimicking a typical BCI interaction, we have been able to identify and isolate a factor inducing non-stationarities with a deep impact on the feature dynamics. The retained information can be utilized for automatically triggered classifier adaptation. And second, we show that pBCIs are indeed capable to enhance common HMS interaction outside the laboratory. With this, we would like to feed back our experiences made with the use of BCIs for HMS. We hope to povide new and useful information about brain dynamics which might be helpful for ongoing research in augmented cognition.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 759–771, 2009.

1 Introduction

Consider the following situation: A user clicks on a file on her computer by mistake. The computer processes the information and starts to open the corresponding application. But this takes some time. She immediately recognizes her mistake and prepares to close the application right after it opens so that she can continue her intended task. She feels distracted and helpless and her feelings are accompanied by facial expressions and inner thoughts. How would it be if the technical system could understand her mistake by analyzing selective information of her, the user? Like humans do in face-to face communication, the system would recognize the mistake almost as soon as one did it and adapt accordingly. This article describes our approach to integrate implicit information into human-machine interaction.

Human-machine systems are defined as systems that integrate goal-oriented cooperation of human beings and technical systems to fulfill a specific task. Regarding this definition, the user has to communicate with the system to execute the given task. To do so the human has to interact with the system using some interface. In most of today's human-machine systems, this interface requires the user to explicitly state his or her demands and wishes to the technical system. This implies that once the command is set by the user the technical system starts to process the command. The system is not able to detect to what extend an action actually concurs to the users' wishes because the system does not have any means to get a real-time estimation of the situation and the user. To interrupt, or end an erroneous action, the user can only issue another command to the system. Current human-machine systems lack the ability to detect implicit information as humans normally do in human face-to-face interaction.

Human-human interaction contains both explicit and implicit information. Humans use both kinds of information to analyze a situation and this way, two-way adaptation of behavior is possible, information gets exchanged and the behavior in a group is regulated. In this context we define explicit interaction as a conscious action to exchange information, for example language and script. Implicit interaction can be defined as an unconscious action that is integrated in another action, for example mimic and gesture. Because of the complexity of humans, both kinds of information have to be analyzed and interpreted to judge a situation and to choose an adequate action. Implicit interaction helps humans to understand situations and behavior and thereby helps to communicate and cooperate with a wide range of different human beings in an efficient and effective way.

For the development of new implicit information channels between human and machine, accessing the human brain seems to be a very promising approach. Methods from the field of Brain-Computer Interfaces (BCI), based on statistical machine learning, have proven to be very successful for achieving this goal. A BCI allows for direct access to cognitive states of the user, without the necessity of any activity outside the central nervous system. The first idea of using brain activity directly for communication comes from Vidal in the early 70s. First steps into the field of application have been realized by Birbaumer and Wolpaw

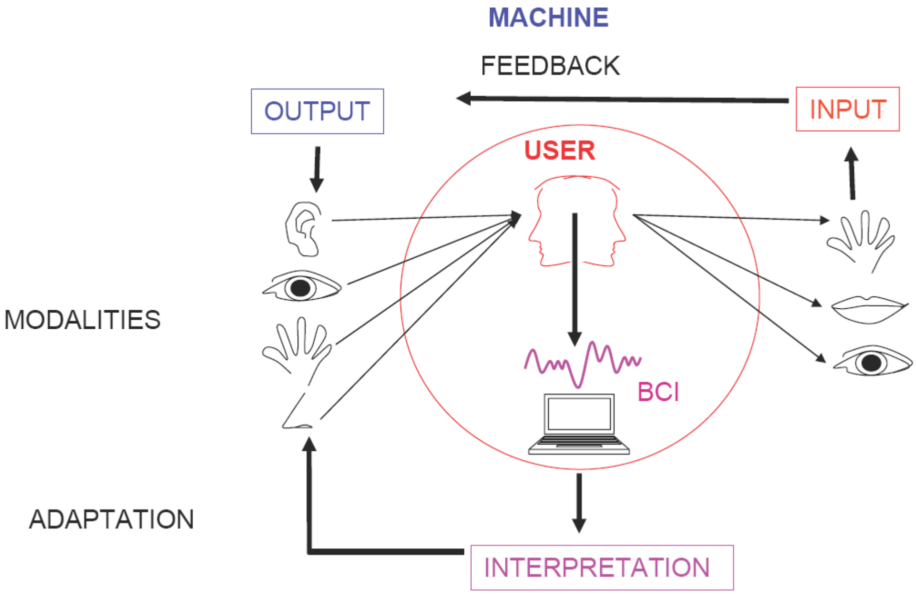


Fig. 1. Brain-Computer Interaction - Feedback cycle within the human-machine system augmented by a BCI-based interpretation input. User states hardly inferable from exogenous factors are estimated by the BCI and fed back into the technical part of the system. The states can be handled as explicit or implicit commands.

in the 80s and 90s by building up support systems for patients suffering from amyotrophic lateral sclerosis (ALS). The introduction of methods from statistical machine learning to BCI had a deep impact on classification accuracy - it allowed for transferring the learning effort from the human being to the machine. Hence, it minimized the effort needed to build up the skill of controlling a BCI system. Additionally, it enabled other fields of research to adapt methods from BCI research for their own purposes. When the focus is laid on the applicability of BCI while interacting in typical human-machine system environments we suggest to use the term Brain-Computer Interaction instead of Brain-Computer Interface.

Team PhyPA, the workgroup for physiological parameters at the department of Human-Machine Systems of the Technical University of Berlin, aims at the combining the technologies from Brain-Computer Interfacing (BCI)[8, 9, 1] and those from the context of Human-Machine Systems (HMS). Therefore, an interdisciplinary team of mathematicians, psychologists and engineers works on currently seven projects investigating non-stationarities, efficiency and general applicability of feature extraction methods, single trial detection of motor and non-motor patterns e.g. error-responses and defining support systems enhancing HMS. As we focus on the applicability of BCI while interacting in typical HMS environments we augment the field of Brain-Computer Interfaces to the field of *Brain-Computer Interaction* (see fig.1). Therefore, we have developed several tools allowing us to detect intended and non-intended user states and integrate them into existing and new HMS. We categorize the methods derived from BCI research into *active* and *reactive*. By the term active BCI (aBCI) we denote BCIs which utilize brain activity of direct correlates of intended actions as input. This includes the detection of motor imagery or execution as well as the control over slow cortical potentials. A reactive BCI (rBCI) is still controlled via intended actions. In contrast to the aBCI features are not derived from direct correlates to these actions, but from cognitive *reactions* on exogenic stimuli, as e.g. in the P300 speller. The rBCI features seem to be more robust in general. This might be due to the fact, that they usually depend on automatic processes of cognition which are not as easily modulated by conscious processes. According to this line of thought [3, 4], we now define *passive* BCI (pBCI). pBCIs are based not on intended actions of the user, but instead on reactive states of the user's cognition automatically induced while interacting in the surrounding system. Hence, the underlying features used by pBCIs are mostly independent of the primary mode of interaction within an HMS, be it BCI based or not.

2 Methods

2.1 Specifications of Our BCI System and Experimental Design

2.1.1 BCI Structure

The PhyPA-BCI is a distributed system consisting of a Recording Unit, Feedback Unit, Processing Unit and the EEG Equipment, managed by a C++ based scriptable control software (MiniGUI). Signal processing, inference, etc. is done using our MATLAB based BCI toolbox (PhyPA toolbox) and the user interface is provided by a C++ based feedback generator. All parts of the system are loosely coupled to each other and can run of separate machines. Consequently, a highly controlled information flow is ensured. Stimuli and feedback can be generated for visual, auditory and tactile output, and input can be received via several modes such as keyboard and game controllers.

2.1.2 Recording

The EEG system has 32 channels of Ag/AgCl conventional (EasyCap) or impedance optimized (ActiCap) electrodes. Signals are amplified by a BrainAmp DC system and recorded by the BrainVision Recorder (BrainProducts). The electrodes are distributed on standard 10/20 based caps with 128 positions. Depending on the type of experiment they are placed over according parts of the cortex. Additionally, we record electrooculogram (EOG) for controlling feedback-induced correlated eye movements, and electromyogram (EMG) on the relevant limbs, for protocolling correlated movements. Both are bipolarly multiplexed by a BrainAmp (ExG) system and derived with Ag/AgCl electrodes. In order to retain information on exogenic factors, we also record ambient temperature and noise level within the laboratory.

2.1.3 Analyses

For enabling the technical system to detect different brain-states while they apear, a classifier has to be trained. This is done by collecting sterotype data (training data) for the to be distinguished classes and extract their statistical properties in offline analyses. The offline evaluation followes a method known from machine learning as cross-validation (CV) [10], in which a classifier is repeatedly trained and tested on disjoint trial sets. Out of the several data partitioning variants, such as randomized CV, blockwise or chronological CV, we chose randomized 10x repeated 10-fold CV as it is the most widely used method in the current BCI field. For the estimation of meta-parameters, e.g. for regularization, we performed a 10x repeated 10x10-fold nested CV.

Common to all examined BCI algorithms are the following steps: First, the raw EEG data for a trial is preprocessed in a strictly causal way. Then, a short feature vector is extracted from the preprocessed data. Finally, a classifier is employed to map from the feature vector to a binary decision value. Implementations closely follow their cited reference descriptions. The CSP and SpecCSP methods were successfully applied in several online control sessions in our lab e.g. via the well-known basket paradigm (see video on www.phypa.org). The SCP algorithm implementation was also validated in an upcoming online study.

The EEG features that enable the discrimination fall into two categories: Event-Related Potentials (ERPs) features, and Event-Related (De)Synchronization [11] (ERD/ERS, henceforth called ERD for brevity) features.

Classifiers are chosen from several linear (LDA, RDA, SVM) and non-linear (kernel SVM, rQDA, GMM) methods. In all analyses presented subsequently, (regularized) LDA was the best performing classifier and was therefore selected. Significance statements are substantiated by standard T-Tests and F-Tests without assumptions on the type of underlying distributions.

Table 1. Categorization of BCI Systems and their fields of of application

Type of BCI	Based on features from	Used for
Active	intended generated cognition	direct control
Reactive	unintended changes in cognition by voluntary focussing on exogenic stimuli	direct control, brain switch
Passive	unintended changes in cognition induced by common interaction	supporting systems, user-state detection

2.1.4 Experimental Design

The stimulus presentation in calibration phases before online feedback is designed for providing high control over exogenous and correlating factors besides the one of interest. This control is relaxed in certain online feedback sessions to allow for a more realistic mode of interaction. Notice, that this decrement of control might allow for a higher number of artifacts but does *decrease* the signal to noise ratio. Subjects have been introduced to the main factor of investigation by an instructor. Experimental tasks have been presented in a standardized way on the screen of the Feedback Unit. The course of the experiments contained several breaks for relaxation and recovering of the subjects. Subjects gave information on their overall state and their impressions on different blocks of the experiment by answering questionnaires. All subjects are from age 18 to 45 with german as primary language. All groups of subjects are of balanced or selected sex. After all sessions the subject has been paid (20 Euro).

2.2 The RLR Paradigm and Its Directed Restriction, the RLR-Game

In the Rotation-Left-Right (RLR) paradigm [12] a stimulus on a starting position has to be rotated left- or rightwise (by a left or right key press) until it corresponds to a given target figure. The stimulus is either the letter "L" or "R", indicating the direction of rotation. While the colour of the stimulus is grey, it can not be rotated. However, every 1000 ms it changes into one of three colours, indicating A) the possibility to be rotated by a keypress and B) the degree of rotation. If the stimulus lights up in red, the stimulus will rotate 90 degree, if it is yellow 60 degree and if it is green 30 degree. Please notice, that each rotation has to be triggered, which only can be done once per colour change. The subject has to build up an efficient strategy for reaching the target: to rotate the starting stimulus as fast as possible on the target stimulus without rotating too far. A derivate of the RLR paradigm is the RLR-Game, defined in two modes (see figure 2): The first was restricted to standard states and the second with additional error states. The standard states are restricted to the colours green and red. The mapping of angles in the error states is directed downwards, hence an error induces a smaller angle of rotation than indicated by the colour. Goal of the game is to reach the target stimuli as fast as possible. Two players can play against each other. Their performance is measured and fed back in points. A player get a point when hitting the target earlier than his opponent.

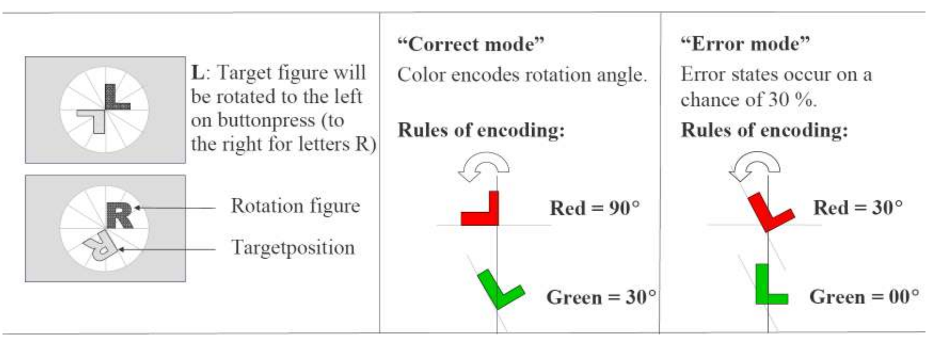


Fig. 2. States of the RLR-Game, rules for rotation in correct and error mode

3 Experimentel Setups

3.1 A pBCI for Retaining Interaction Relevant Endogenous Information

3.1.1 Motivation

Shifting BCI applications from laboratory environment to interactive scenarios enforces losing the control over most of the interfering factors. Hence, one faces problems connected to the interaction between man and machine. The use of pBCI might give insights into the correlation between mental states and system states which are hard to infer from exogenous factors. This study investigated the factor of perceived Loss of Controllability (LoC). By controllability we refer to the perceived control the user has over a feedback device. Noticing classification errors the user is aimed at regaining control over the machine. Perceived LoC could cause a change in the mental state of the subject, and therefore have an impact on feature space. The idea of perceived controllability has been stated before, for it refers directly to the classification accuracy of a BCI system [13, 14]. Data based methods face the problem of complexity of online BCI systems, consisting of two strongly interacting components, namely the user and the machine, which creates a closed feedback loop. Each of the interacting systems has to optimize for the same goal - hence ideally adaptations of both systems should converge. But in the other case it could be, that both systems diverge from one another, as demonstrated here. LoC is a result of the static translation algorithms confronted with the variable brain trying to optimize during phases of classifier errors. LoC can be defined as the classifier output being inconsistent with expected feedback. Hence, with the RLR paradigm we manipulated the rate by which the user was able to predict the feedback under controlled conditions, by artificially inserting machine errors. The study was held in offline mode, to ensure the control for possible intervening factors and to avoid phases of loss of control over the feedback device, as they occur in online sessions. This study has found evidence for a crucial factor causing nonstationarities, which indicate the percieved LoC. This BCI related problem can be seen from a more abstract perspective. The identification of the correlated non-stationarities gives insight into the feeling of losing the control over a device. This factor can be utilized to enhance human-machine interaction in a special way. It allows for access to a user state, which is hardly accessible by known apllicable methods. This user state is *hidden* from the system standard input or sensing methods. Here it is accessible by taking physiological parameters into account.

3.1.2 Factor of Investigation

One class of problems is that of non-stationarities resulting from shifts in cognitive states, which have not been represented in the data of the calibration phase. These might be induced by changes in the mode of interaction or mental processing of exogenic factors. As stated by Dornhege, Shenoy and Krauledat (see `www.bbci.de`) the loss of controllability (LoC) might be one of these factors.

3.1.3 Experimental Design

By utilizing the RLR paradigm we have been able to artificially induce phases of reduced controlability (BUc, see fig. 3) in experiments with 24 subjects by permuting the mapping between colours and angles of rotation. We tracked features representing the primary mode of interaction, pressing a key, in the EEG data. One representing the event-related desynchronization (ERD) and one representing a slow cortical potential (SCP) prior to the movement. Details on this study can be found in [12].

The CSP Algorithm: Feature extractor for ERDs and SCPs used here is the Common Spatial Patterns for SCP (CSPfSCP) algorithm [16]. CSP aims to find linear combinations (patterns) of EEG channels such that the deflection of each trial projected according to these patterns is most discriminative (i.e., differs maximally between the two classes).

The SpecCSP algorithm: For the extraction of ERD features we used Spectrally Weighted CSP (SpecCSP) [17]. SpecCSP iteratively alternates between optimizing spatial and the spectral criteria. This way, the algorithm calculates a set of custom spatial projection together with a set of custom frequency filters.

Classifiers: The classifier we used is Linear Discriminant Analysis (LDA), which is optimal as long two requirements are met: noise projected must be Gaussian and uncorrelated to the class membership.

3.2 Dependent Measures of Nonstationarities

To assess the impact of LoC on the classifier's performance, we calculated pseudo online classification rates (POC) over time. POC rates were calculated by offline analysis serving as estimation for online classification results. POC was determined as following: The appropriate CSP derivate was used, with a time window of 300ms, six patterns and a band-pass filter of [7-30] Hz. A classifier was trained on the initial LR block. Then, this classifier was applied to every key press, which happened over the course of the main experiment (i.e. blocks A1, A2 and B). An average of approx. 100 gradual classifier outputs in a one-second window before each key press was averaged and taken as the classifier's decision for this key press. The sign of this decision value (by default, left keys, on average, assigned -1, right keys +1) was remapped according to the key actually pressed, such that correct decisions were assigned positive values and wrong decisions were assigned negative values. By this, we got a real number for each key that was pressed by the subject. We also calculated the Kullback- Leibler divergence (KLD) (for details see [14]) of the classifier's feature distributions. All measures were calculated relative to the training data's distribution of the initial calibration measurement, the LR-block. We used the KLD to measure the divergence of the CSP feature distributions as they build up over the course of the main experiment. Note that these KLD results have been determined in a classical offline fashion, i.e. one for each key press. We have found a significant (¡0.05) correlation of the of all KLD and POC values over all 22 subjects and a significant (¡0.05) difference in KLD

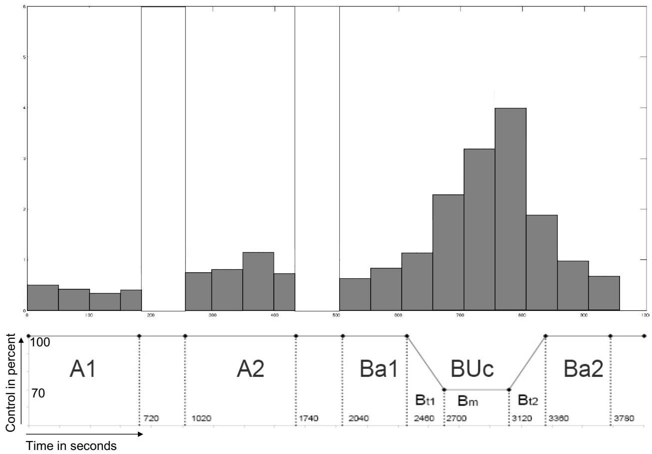


Fig. 3. Grand average of the KL divergence in CSP features through time

between A1+A2 and BUc in SpecCSP features, but none of this in CSPfSCP features. The accumulated course of the SpecCSP-KLD is shown in figure 3.

3.3 Applicability of a pBCI for Enhancement of Efficiency in HMS

3.3.1 Motivation

Errors in communication are highly relevant factors regarding the efficiency of HMS. Especially in automated adaptation of the machine to the interaction mode of the user [18]. A wrong decision induces effects of surprise and frustration and in this respect, adaptation reduces the performance and the safety in HMS [19]. Additionally it triggers a correction action which enforces a shift in the intention focus of the user. According to this it reduces the overall acceptance of the adaptation and of the whole system.

3.3.2 Factor of Investigation

In this study we have shown that pBCIs are capable of enhancing such an adaptation. For this we have designed the RLR-Game which mimics the interaction in an HMS and allows for modelling an unexpected and negative effect, the error states. While this game is based on common interaction channels we have added a secondary and passive BCI channel capable of correcting the effect of error states. This correction was triggered by an event related potential reflecting the mental processing of an error trial. If it is correctly detected by the pBCI during an error trial, the rotation angle was set to the correct mapping. In case of a false positive the angle was reduced to that of an error state. Hence, each correct detection of an error speeds the player up and a false positive slows him down. Therefore, if the classifier works properly, it will enhance the performance of the player and it will reduce it otherwise. This approach utilizes BCI methodology for instantaneous adaption to the user. And also, like in the LoC study it would give access to hidden user states. But the information gained is used as a passive and secondary input methods, which would allow for a more natural and efficient interaction between men and machine.

3.3.3 Experimental Design

For keeping the environment as realistic as possible, we have chosen the Open House of the TU Berlin (LNdW 2007) as the setting. Four times two different players from the audience played the game against each other. Each pair played three sessions of 50 trials. One for user training, without error states. In the second session we introduced the error trials. The automatic adaptation has been applied in the last session, only for one player.

3.3.4 Features

Features have been extracted by a derivate of the pattern matching algorithm ([20]) extended for detection of several extrema of SCPs. 600 ms of data after the rotation have been selected.

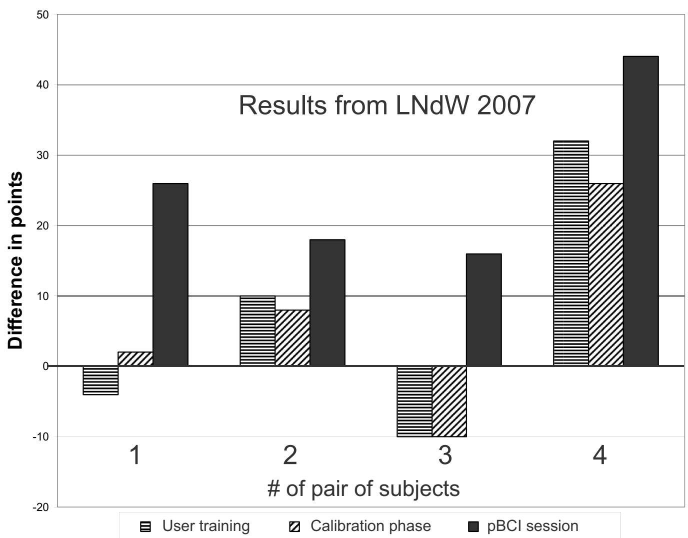


Fig. 4. Differences of points from two opponents playing the RLR-Game at the Open House 2007

4 Results

The results of the LoC study (fig. 3) show that in phases with full control (A1, A2, Ba1, Ba2) the variance of the averaged Kullback-Leibler divergence (KLD) of both features is bounded. In contrast, the phases of reduced controlability (BUc) shows an significant increase of the KLD for ERD based features. Hence, the

KLD of these features is a measure representing the perception of controlability. Figure 4 shows the results from the sessions from the open house of the TU Berlin 2007. In the third session one player has been supported by the pBCI. While the points have been equally distributed between session 1 and 2, the performance of all pBCI supported players has been increased significantly.

5 Discussion

Here we gave examples of two types of pBCIs. One establishing an information flow from the human brain to the HMS reflecting user states correlated to current modes of interaction. The other one extracting the actual interpretation of dedicated system states from the users cognition. Both can be applied in the context of BCI for enhancing classification accuracy. First, for an automated adaptation of the classifier and second, for correcting machine errors as proposed in [20, 21]. Also, for application in the field of HMS, it provides information about the user, which can only hardly be inferred by typical information channels in HMS. Especially the idea of utilizing the human brain as sensory for the subjective interpretation of current states within the HMS seems to be very promising. These studies are hopefuly a starting point for a whole series of new approaches. Currently we are investigating pBCIs for detection of mental workload, cognitive interpretation of the perception of human movements [22] and information on driver intentions. Please, see `www.phypa.org` for details.

6 Conclusion

Our experiences with pBCIs show, that these enable new channels of information within the interaction between man and machine. Here we show two examples of new approaches for enhancing human-machine interaction. Both are based on interpretation of brainwaves by BCI methodology, but describe almost orthogonal ways of interacting with a technical system. First, within the LoC study, we show that a BCI equipped system is indeed capable of detecting user states which are hidden from exogenous interpretation. Besides the LoC there are much more of these hidden user states, for example like mental workload, attention focus or arousal. All of these allow for changing the actual system mode and adapt the system to the user in a broader sense. Hopefully, there will be solutions for providing this information to HMS in near future. The second approach describes a new way for spontaneous and immediate adaptation of the system to the needs of the user, in an event-related fashion. Here the system also gains access to hidden user states, which are hardly detectable by exogenous methods. But here the input drawn from the EEG is a new passive way of direct communication between human and machine. These approaches can be utilized for both BCI and HMS research. Additionally it seems to be very fruitful to exchange experiences between these two fields of research, which will hopefully will be done extensively in the near future.

References

[1] Blankertz, B., Curio, G., Mueller, K.-R.: Classifying single trial eeg: Towards brain computer interfacing. In: Advances in Neural Inf. Proc. Systems (NIPS 2001), January 2002, vol. 14, pp. 157–164 (2002)

[2] Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.-R., Curio, G.: The non-invasive berlin brain-computer interface: Fast acquisition of effective performance in untrained subjects. NeuroImage 37, 539–550 (2007)

[3] Zander, T.O., Kothe, C., Jatzev, S., Dashuber, R., Welke, S., de Fillippis, M., Roetting, M.: Team phypa: Developing applications for brain-computer interaction. In: Proceedings of the CHI-Conference, Florence, Italy (2008)

[4] Cutrell, E., Tan, D.: BCI for passive input in HCI. In: Proceedings of the CHI-Conference, Florence, Italy (2008)

[5] Zander, T.O., Kothe, C., Jatzev, S., Luz, M., Mann, A.: Das phypa-bci: Ein brain-computer-interface als kognitive schnittstelle in der mensch-maschine-interaktion. Prospektive Gestaltung von Mensch-Technik-Interaktion 13, 183–185 (2007)

[6] Guger, C., Edlinger, G., Harkam, W., Niedermayer, I., Pfurtscheller, G., Oeg, G.T., Graz, A.: How many people are able to operate an eeg-based brain-computer interface (bci)? IEEE Transactions on Neural Systems and Rehabilitation Engineering 11, 145–147 (2003)

[7] Blankertz, B., Dornhege, G., Krauledat, M., Muller, K.R., Curio, G.: The berlin brain-computer interface: Report from the feedback sessions. Technical report, Technical Report 1, Fraunhofer FIRST (2005)

[8] Vidal, J.J.: Toward direct brain-computer communication. Annual Reviews in Biophysics and Bioengineering 2, 157–180 (1973)

[9] Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain-computer interfaces for communication and control. Clinical Neurophysiology 113, 767–791 (2002)

[10] Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. Wiley-Interscience, Hoboken (2000)

[11] Pfurtscheller, G., Lopes: Event-related eeg/meg synchronization and desynchronization: basic principles. Clinical Neurophysiology 110(11), 1842–1857 (1999)

[12] Jatzev, S., Zander, T.O., DeFlilippis, M., Kothe, C., Welke, S., Roetting, M.: Examining causes for non-stationarities: The loss of controllabilityi s a factor which induces non-stationarities. In: Proc. of the 4th Int. BCI Workshop & Training Course, vol. (4) (2008)

[13] Dornhege, G.: Increasing Information Transfer Rates for Brain-Computer Interfacing, PhD thesis. University of Potsdam (2006)

[14] Shenoy, P., Krauledat, M., Blankertz, B., Rajesh, P., Rao, N., Müller, K.-R.: Towards adaptive classification for bci. Journal of Neural Engineering 3, R13–R23 (2006)

[15] Cui, R.Q., Huter, D., Lang, W., Deecke, L.: Neuroimage of voluntary movement: topography of the Bereitschaftspotential,a 64-channel DC current source density study. Neuroimage 9, 124–134 (1999)

[16] Dornhege, G., Blankertz, B., Curio, G.: Speeding up classification of multi-channel brain-computer interfaces: Common spatial patterns for slow cortical potentials. In: Proceedings of the 1st International IEEE EMBS Conference on Neural Engineering. Capri 2003, January 2003, pp. 591–594 (2003)

[17] Tomioka, R., Dornhege, Aihara, G., Aihara, K.K., Müller, K.-R.: An iterative algorithm for spatio-temporal filter optimization. In: Brain Computer Interface Workshop and Training Course, vol. 3, pp. 22–23 (2006)

[18] Parasuraman, R., Sheridan, T.B., Wickens, C.D.: A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man and Cybernetics, Part A 30, 286–297 (2000)

[19] Sarter, N.B., Woods, D.D., Billings, C.E.: Automation surprises. Handbook of Human Factors and Ergonomics 2, 1926–1943 (1997)

[20] Blankertz, B., Dornhege, G., Schafer, C., Krepki, R., Kohlmorgen, J., Muller, K.R., Kunzmann, V., Losch, F., Curio, G.: Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial eeg analysis. IEEE Transactions on Neural Systems and Rehabilitation Engineering 11, 127–131 (2003)

[21] Ferrez, P.W., Millan, J.R.: Error-related EEG potentials in brain-computer interfaces. In: Towards Brain-Computer Interfacing, pp. 291–301. The MIT Press, Cambridge (2007)

[22] Gärtner, M., Klister, T., Zander, T.O.: Classifying the observation of feasible and unfeasible human motion. In: Proc. of the 4th Int. BCI Workshop & Training Course, vol. (4) (2008)

Augmented Cognition as Rehabilitation: Facilitating Neuroplasticity?

Michael Feuerstein[1,2], Gina Luff[3], Mark Peugeot[1], Miki Moskowitz[1], and Briana Todd[1]

[1] Uniformed Services University of the Health Sciences, Bethesda, MD
[2] Georgetown University Medical Center, Washington, DC
[3] American University, Washington, DC
{mfeuerstein,gluff,Mark.Peugeot,Michal.Moskowitz,btodd}@usuhs.mil

Abstract. Different types of brain injury are associated with deficits in working memory, executive functioning, and information processing speed, which can impact performance at work. Augmented Cognition (AugCog), a technology developed to improve human performance in complex tasks, may have potential for optimizing cognitive functioning in the context of work for those with mild to moderate cognitive deficits. AugCog is a way to accommodate or augment function thus improving the performance of the operator. This approach may facilitate neuroplasticity that can occur following injury to the brain. The authors will provide the rationale, operational structure, and potential application to occupational rehabilitation.

Keywords: Augmented cognition, occupational rehabilitation, cognitive limitations, mild brain injury.

1 Introduction

Injury to the human brain can occur following various types of exposures. Automobile accidents, sports related injuries, and cytotoxic chemicals including those used to treat many forms of cancer have all been associated with sequelae that reflect some type of mild brain injury. There has been recent concern with the rise in mild traumatic brain injury (mTBI), the signature injury of current US military conflicts in Iraq and Afghanistan, due to the blast effects of Improvised Explosive Devices (IEDs) [1]. A significant number of cancer survivors also report cognitive problems that resemble mild forms of brain injury. These problems are thought to be associated with exposure to chemotherapy and/or radiation [2, 3, 4, 5].

Trauma to the central nervous system is associated with limitations in many areas of cognitive function, including difficulties in working memory, executive function, and attention [6]. These cognitive functions are essential for optimal performance at

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 775–781, 2009.

work [7]. Accordingly, many of those with mTBI experience difficulties maximizing their work abilities, returning to work, or remaining at work [8, 9].

Cognitive neuroscience and human factors research and practice have evolved to the point where their integration has created a technology designed to optimize human function in complex work environments [10, 11]. The development of augmented cognition represents one such approach to enhancing human performance during military operations [10]. The theory and techniques of augmented cognition may be extended to civilian work environments in order to improve cognitive and behavioral performance among those working or attempting to work with cognitive limitations associated with a mild brain injury. The application of augmented cognition to the modern workplace is especially relevant because much of today's work involves human-computer interaction.

The present paper discusses the assumptions forming the foundation of augmented cognition as a means for enhancing cognitive related functional outcomes in those with mild cognitive limitations secondary to brain injury. We first consider a few of the similarities in the types of limitations in cognitive function in different mild brain injuries. Next are the assumptions that form a basis for applying the augmented cognition technology to occupational rehabilitation. A description of a potential work rehabilitation system based on augmented cognition technology will follow. Finally, potential applications within a work rehabilitation context are discussed.

1.1 Cognitive Limitations and Mild Injury to the Brain That Can Impact Work Ability

Regardless of its etiology, while some functional loss is related to damage in focal brain areas, mild trauma to the human brain (e.g., classic mTBI, cancer treatment) is often associated with a pattern of generic changes in cognitive functions, such as information processing speed [12] and working memory [12, 13, 14, 15, 16]. The specific problems observed in those with some injury to the brain typically also include disruption of executive function [7], which is characterized by a difficulty in coordinating tasks, multitasking, problem solving, attention, and planning. Additionally, when the exposure is mild and the person's capabilities were relatively intact prior to the trauma, the person is typically aware of such changes when confronted with tasks that involve these cognitive functions. Some have described the changes as a "premature aging," where a person's cognitive abilities were intact prior to the brain insult, but following the injury, cognitive operations were not as fluid, coordinated, rapid, or error free. These cognitive-motor changes can occur along with certain emotional changes such as depression [17], heightened levels of stress including some elements of Post-Traumatic Stress Disorder, and what is called sub-clinical depression, where a person is not seriously depressed but has symptoms of fatigue, negative mood, sleeplessness, etc. [18, 19]. These factors can interact in a manner that exacerbates cognitive limitations and thus impact the ability to perform optimally at work [20].

1.2 Basic Assumptions Underlying AugCog Rehab Approach

As with any new rehabilitation approach, certain assumptions are necessary in order to define the process and its potential outcomes. The following assumptions, supported by varying levels of empirical evidence, are presented in order to justify the potential application of augmented cognition for occupational rehabilitation.

1. The augmented cognition approach has demonstrated some success in improving human performance in complex work situations [11].
2. Changes in various neurophysiological signals such as EP and EEG synchronization have been observed in response to cognitive demands in many different types of mild brain injury [4, 6, 21, 22].
3. It is possible to identify neurophysiological correlates of cognitive states referred to as cognitive signatures, which are unique to the individual and can be used as markers of sub-optimal performance [23].
4. These cognitive signatures can trigger various mitigation strategies that will facilitate optimization of performance [24].
5. The mitigation strategy and improvement in performance can in turn alter brain function (neurophysiology and perhaps blood flow) from a state of a “sub-optimal” neural behavioral response to a more “optimal” brain-behavior state [24].
6. Research on brain plasticity has shown that repetitive training in specific tasks, which may involve use of prostheses, can facilitate neural changes [25, 26]. These fundamental concepts can be applied to the context of work. It is expected that the augmented cognition process, when operating repetitively with a realistic set of work tasks that engage the cognitive operations that are problematic for the affected worker, can improve both the pattern of brain activity/function and behavioral correlates involved in improved work performance. The generalization of more functional neural-behavioral patterns over time would also be related to improved work performance over time.
7. The potential underlying mechanism of this relearning is the facilitation of brain plasticity/remodeling occurring over time [26].

2 Elements of AugCog Occupational Rehabilitation System (Fig 1)

The AugCog Occupational Rehab System involves the monitoring of neural and autonomic signals, superficial blood flow on the surface of the brain, and behavioral performance in individually relevant work tasks reflecting cognitive-behavioral performance. An executive processor then analyzes this information and triggers specific mitigations, which in turn help to improve the performance of the affected worker. This can facilitate recovery of brain and behavioral functioning.

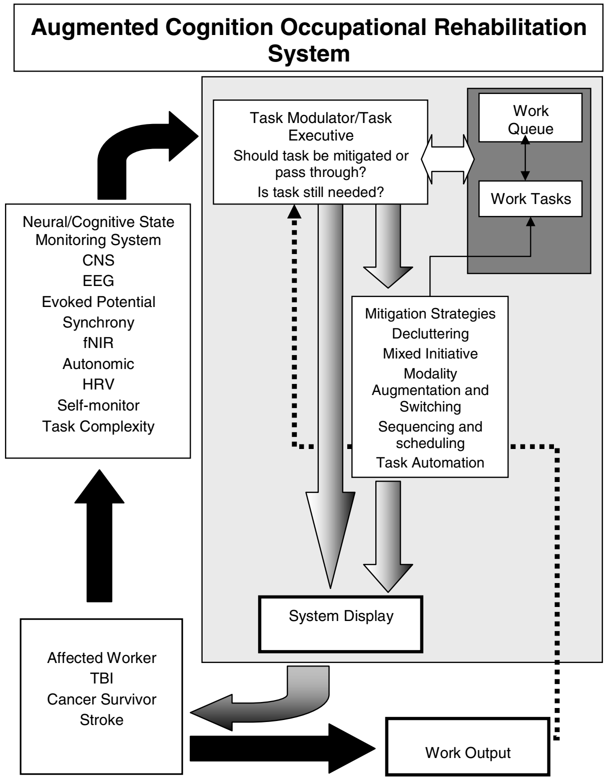


Fig. 1. General illustration of the elements of the system and their interactions. This proposed system was refined through the input of Honeywell (Trish Ververs and Santosh Mathan), Design Interactive (Kay Stanney), and Archionoetics (Hunter Downs).

3 Applications

As the population continues to age and societal regulations and personal preferences extend employment for years, there will be a need to improve cognitive work capacity in addition to efforts to engineer out the problems faced by many workers with and without illnesses that impact cognitive performance. The rehabilitation system proposed, if effective, will be able to assist those who are experiencing cognitive limitations that interfere with their work. This would not only include the types of brain injuries discussed above, but it also may be applicable to age-related change in cognition and work performance.

Another viable application of this methodology would be to create a miniaturized version of the system for individuals to use as an ongoing work accommodation tool or neural prosthesis. The system, which would interface with any computer platform, would continuously and non-invasively monitor neural cognitive and behavioral performance. By generating appropriate mitigations in milliseconds, the system would assist the affected worker in maintaining optimal work performance. Such a neural prosthesis would not require long term change in brain function; rather, it would engineer the work process so that neuro-cognitive-behavioral changes that accompany performance changes can be rapidly mitigated. The system would continuously and rapidly ensure that the affected worker's productivity is maintained while the worker would simply continue his or her task uninterrupted.

4 Conclusion

Once the rehabilitation system is developed, it must be tested to determine its ability to perform as expected. The ability to generate any change in work performance also needs to be evaluated. Changes in neural activity underlying optimization of work productivity must also be determined. Randomized controlled trials must be conducted in order to evaluate the effects of the approach on work performance and work retention. If the approach is demonstrated to be effective, studies of markers of plasticity would also be useful in order to reveal potential mechanisms of change. Much work needs to be done. However, this proposed approach derived from the HCI field holds much potential in an area lacking effective occupational rehabilitation.

Acknowledgements

The authors wish to thank Trish Ververs, Santosh Mathan, Kay Stanney, and Hunter Downs for their valuable input on the elements of the system. Without their contributions and that of Murray Cooper at Honeywell, we would never be this far along. We are also deeply grateful for the early advice and encouragement of Dylan Schmorrow and Joseph Cohn.

References

1. Tanielian, T., Jaycox, L.H. (eds.): Invisible Wounds of War: Psychological and Cognitive Injuries, Their Consequences, & Services to Assist Recovery. RAND Corporation, Santa Monica (2008)
2. Kreukels, B.P.C., Schagen, S.B., Ridderinkhof, K.R., Boogerd, W., Hamburger, H.L., van Dam, F.S.A.M.: Electrophysiological Correlates of Information Processing in Breast-Cancer Patients Treated with Adjuvant Chemotherapy. Breast Cancer Research and Treatment 94, 53–61 (2005)
3. Kreukels, B.P.C., Hamburger, H.L., de Ruiter, M.B., van Dam, F.S.A.M., Ridderinkhof, K.R., Boogerd, W., Schagen, S.B.: ERP Amplitude and Latency in Breast Cancer Survivors Treated with Adjuvant Chemotherapy. Clinical Neurophysiology 119, 533–541 (2008)
4. Remer, S., Murphy, M.E.: The Challenges of Long-Term Treatment Outcomes in Adults with Malignant Gliomas. Clinical Journal of Oncology Nursing 8, 368–376 (2004)
5. Schagen, S.B., Hamburger, H.L., Muller, M.J., Boogerd, W., van Dam, F.S.A.M.: Neurophysiological Evaluation of Late Effects of Adjuvant High-Dose Chemotherapy on Cognitive Function. Journal of Neuro-Oncology 51, 159–165 (2001)

6. Serino, A., Ciaramelli, E., Di Santantonio, A., Malagu, S., Servadei, F., Ladavas, E.: Central Executive System Impairment in Traumatic Brain Injury. Brain Injury 20, 23–32 (2006)
7. Grafman, J.: Executive functions. In: Parasuraman, R., Rizzo, M. (eds.) Neuroergonomics: The Brain at Work, pp. 159–177. Oxford University Press, Oxford (2008)
8. Stulemeijer, M., Van Der Werf, S.P., Jacobs, B., Biert, J., Van Vugt, A.B., Brauer, J.M.P., Vos, P.E.: Impact of Additional Extracranial Injuries on Outcome after Mild Traumatic Brain Injury. Journal of Neurotrauma 23(10), 1561–1569 (2006)
9. Binder, L.M., Rohling, M.K., Larrabee, G.J.: A review of mild head trauma, part I: Meta-Analytic Review of Neuropsychological Studies. Journal of Clinical and Experimental Neuropsychology 19(3), 421–423 (1997)
10. Parasuraman, R., Rizzo, M.: Introduction to Neuroergonomics. In: Parasuraman, R., Rizzo, M. (eds.) Neuroergonomics: The Brain at Work, pp. 3–12. Oxford University Press, Oxford (2008)
11. Stanney, K.M., Schmorrow, D.D., Johnston, M., Fuchs, S., Jones, D., Hale, K., Ahmad, A., Young, P.: Augmented Cognition. Reviews of Human Factors and Ergonomics 5 (in press)
12. Schilder, C.M., Eggens, P.C., Seynaeve, C., Linn, S.C., Boogerd, W., Gundy, C.M., Beex, L.V., van Dam, F.S., Schagen, S.B.: Neuropsychological Functioning in Postmenopausal Breast Cancer Patients Treated with Tamoxifen or Exemestane After AC-Chemotherapy: Cross-Sectional Findings from the Neuropsychological TEAM-Side Study. Acta Oncologia 48, 76–85 (2009)
13. Castellon, S.A., Ganz, P.A., Bower, J.E., Petersen, L., Abraham, L., Greendale, G.A.: Neurocognitive Performance in Breast Cancer Survivors Exposed to Adjuvant Chemotherapy and Tamoxifen. Journal of Clinical and Experimental Neuropsychology 26(7), 955–969 (2004)
14. Jenkins, V., Shilling, V., Fallowfield, L., Howell, A., Hutton, S.: Does Hormone Therapy for the Treatment of Breast Cancer Have a Detrimental Effect on Memory and Cognition? A Pilot Study. Psycho-Oncology 13, 61–66 (2004)
15. McHugh, T., Laforce, R., Gallagher, P., Quinn, S., Diggle, P., Buchanan, L.: Natural History of the Long-Term Cognitive, Affective, and Physical Sequlea of Mild Traumatic Brain Injury. Brain and Cognition 60, 209–211 (2006)
16. Lee, H., Wintermark, M., Gean, A.D., Ghajar, J., Manley, G.T., Mukherjee, P.: Focal Lesions in Acute Mild Traumatic Brain Injury and Neurocognitive Outcome: CT Versus 3T MRI. Journal of Neurotrauma 25, 1049–1056 (2008)
17. Reppermund, S., Ising, M., Lucae, S., Zihl, J.: Cognitive Impairment in Unipolar Depression is Persistent and Non-Specific: Further Evidence for the Final Common Pathway Disorder Hypothesis. Psychological Medicine (2008) (epub ahead of print)
18. Hanna-Pladdy, B., Berry, Z.M., Bennett, T., Phillips, H.L., Gouvier, W.D.: Stress as a Diagnostic Challenge for Postconcussive Symptoms: Sequelae of Mild Traumatic Brain Injury or Physiological Stress Response. The Clinical Neuropsychologist 15(3), 289–304 (2001)
19. Hoge, C.W., Terhakopian, A., Castro, C.A., Messer, S.C., Engel, C.C.: Association of Posttraumatic Stress Disorder with Somatic Symptoms, Health Care Visits, and Absenteeism among Iraq War Veterans. American Journal of Psychiatry 164(1), 150–153 (2007)
20. Lupien, S.J., Maheu, F., Tu, M., Fiocco, A., Schramek, T.E.: The Effects of Stress and Stress Hormones on Human Cognition: Implications for the Field of Brain and Cognition. Brain and Cognition 65, 209–237 (2007)

21. Lew, H.L., Poole, J.H., Chiang, J.Y.P., Lee, E.H., Date, E.S., Warden, D.: Event-Related Potential in Facial Affect Recognition: Potential Clinical Utility in Patients with Traumatic Brain Injury. Journal of Rehabilitation Research & Development 42(1), 29–34 (2005)
22. Kreukels, B.P.C., Schagen, S.B., Ridderinkhof, K.R., Boogerd, W., Hamburger, H.L., Muller, M.J., van Dam, F.S.A.M.: Effects of High-Dose and Conventional-Dose Adjuvant Chemotherapy on Long-Term Cognitive Sequelae in Patients with Breast Cancer: An Electrophysiologic Study. Clinical Breast Cancer 7, 67–78 (2006)
23. Mathan, S., Whitlow, S., Feyereisen, T.: WorkSense: Exploring the Feasibility of Human Factors Assessment Using. Electrophysiological Sensors. In: Proceedings of the 4th International Conference of the Augmented Cognition Society, HFES, Santa Monica (2007)
24. Fuchs, S., Hale, K.S., Stanney, K.M., Juhnke, J., Schmorrow, D.: Enhancing Mitigation in Augmented Cognition. Journal of Cognitive Engineering and Decision Making 1(3), 309–326 (2007)
25. Duffau, H.: Brain Plasticity: From Pathophysiological Mechanisms to Therapeutic Applications. Journal of Clinical Neuroscience 13, 885–897 (2006)
26. Taub, E.: Harnessing Brain Plasticity through Behavioral Techniques to Produce New Treatments in Neurorehabilitation. American Psychologist 59, 692–704 (2004)

Embodying Meaning in Bio-cognitive Aid Design

Daniel Garrison[1] and Victoria Garrison[2]

[1] Virginia Polytechnic Institute and State University, College of Engineering, Virginia
djg@vt.edu
George Mason University, School of Information Technology and Engineering, Virginia
dgarriso@gmu.edu
[2] Underwood-Memorial Hospital, Family Medical Center, Woodbury, New Jersey
garrisonv@umhospital.org

Abstract. Through analysis involving components of theory, concept, and semantics, we synthesize a set of insights and relations existing through embodiment that have applicability to the design of cognitive aids. Of note, we explore the areas of embodied interaction, embodied user interfaces, and embodied cognition. Additionally, we demonstrate how meaning is created and distributed in and through the areas. We then apply findings to the development of an embodied user interface developed to cognitively aid in renovation projects.

1 Introduction

As we move from understanding cognition to augmenting it, we find ourselves faced with questions on how to effectively embody cognitive aids within the different environments. The concept of embodiment is far reaching within the field of human-computer interaction (HCI). Areas such as tangible, social, ubiquitous, and mobile computing have all subscribed to the fact that the objects of our daily activities are grounded in the practices of the world. According to Dourish [13], it is not simply that objects exist in a physical reality, but rather that they employ a form of participatory status in the world where they are embedded. These objects, whether they are natural or man-made, computational or of a genetic makeup, all find the completion of their meaning in their involvement with other objects or individuals within a particular time interval. It is this involvement that we must capture to take cognitive augmentation and aids to the next level.

The usefulness and even the reality of an aid depend on their embeddedness. Without a world of involvement, these objects are separate from ideas or concepts, isolated from the human cognition. A realization that a particular aid's use depends on the environment where it is exists allows us to connect to human cognition through the substance of the object. In other words, the actions of the objects within a particular environment give the objects meaning. Meaning then is imparted upon the actions and interactions that make up the environment. This is embodiment.

The paper balance is as follows. Section 2 explores current area work that has implications for cognitive aid design. Section 3 looks at embodiment further, to understand what makes up embodiment based upon related terminologies used in the field. Section 4 introduces an implementation of an embodied aid. Section 5 concludes.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 782–791, 2009.

2 Current Research Insights

So if this is how objects gain meaning, what can fields such as tangible computing, ubiquitous computing, social computing, or even mobile computing teach us when applied to the development of cognitive aids? Each field understands, at some level, that the realization of embodiment is necessary to the successful design of systems.

Tangible computing is an approach by which computation is moved off the desktop and into the hands of individuals for manipulation. It is the realization that for aids to be successful in day-to-day activities, objects containing computation need to be grasped to have value and meaning within the day-to-day activities for which they were designed. This is unlike desktop-based computation where both the size and stationary nature of the computer requires stepping away from most types of cognitive activities in order to focus on the computational system. In other words, traditional desktop-based computation disembodies itself from the stream of activities of the world and therefore loses value and meaning that could be gained through intertwined participatory involvement of an aid.

Ubiquitous computing is an area first coined by Mark Weiser [26]. Similar to tangible computing, ubiquitous computing seeks to move and embody aids in environments of activity. A move to focus less on the computer and more on cognitive activities surrounding the computer is approached by deploying small computational devices throughout the environment. The implication is not only direct participatory involvement in a seated environment, but also that the aid's particular use and meaning would be specific to the environment where they interacted. Thus, just as in tangible computing, we see an embodied component to ubiquitous computing.

For social computing, an understanding of sociology and organizational behavior is foundational to research. In addition to other areas, the work seeks to understand how human language, behavior and organizational structure affect and impact the design of computational devices. Contrary to the design of most desktop-based computer applications today is a resistance to viewing the role of computational aids as one that carries out processes. Rather, understanding exists that computational devices are embodied in communities of practice. These communities of practice go beyond processes, to capture the realization that computation exists in a place [19]. A place is more than the space and the physical properties which traditional processes consider. It realizes that social norms and behaviors also contribute. This is how two identical spaces can exhibit very different activities. These social activities bring a set of skills, methods and experiences that exist between people and between people and objects, whether computational or not. In essence, they define how a community of practice carries out processes. Social computing seeks to "know" the embodied environment where computational aids will sit. This realization allows aid design for specific interactions used in a community, rather than just a space where an aid would be used.

Finally, even mobile computing in moving the computer off the desktop and to the user's hands, has realized that how this movement is done depends on the environment where it will be used. It involves not only understanding how an object will be used, but also how the object gains meaning and value in places of use. No longer is the computing device a separate focus of attention but in mobile computing it can be more a part of the communities of activity and gain meaning through its interaction.

3 Faces of Embodiment

Beyond an understanding of embodiment and its requisite standing in active areas of computing research, insights are also derived from embodiment's notions of interaction, user interfaces, and cognition. As we seek to further understand embodiment and its role in the design of cognitive aids, we look at the relationships that these areas have between them. As such, we answer questions concerning their component nature, sameness, and coexistence.

Embodied Interaction. Earlier when we used phrases like *participatory status*, *involvement*, and *actions* we also were describing an area of embodiment known as embodied interaction. Embodied interaction is concerned with how meaning is developed, agreed upon, and communicated in the introduction of technologies of interaction into communities of practice. This seats embodied interaction within the broader area of embodiment, but focuses attention on computational devices as the object of interest in the community. Wenger [27] notes that these communities of practice and the computational devices in them coexist together. Devices are designed and adapted to communities of practice that exist, but these communities of practice are also shaped by the devices within. This seemingly paradoxical statement is fundamental to understanding the interaction that exists in embedded environments. While designing a cognitive aid for a community of practice is plausible, it is not enough. We must also realize that the involvement of the aiding device or technology within the environment also changes and shapes the community. Communities of practice cannot be understood without considering the computational aids that will be in them, and how people cognitively approach and understand the devices. Cognitive aids cannot be designed until we understand how the community of practice will change with the addition of a cognitive aid.

Understanding the "shaping" effect of embodied interaction was the focus of the work of Williams et al. [28] on SignalPlay. Their approach to ubiquitous computing sought to understand the interactions that take place when embodying computation in an environment. Of interest to embodied interaction was whether an introduction of computational devices into the environment would change the cognitive approaches currently in place. Would extra computational responsiveness help or hinder comprehension at sites of social action and meaning? What they found agrees with our earlier point that computational devices actually shape communities of practice. Users were influenced by augmented computational objects and their interaction with the objects additionally formed a new method of communicating with each other. Additionally, users found an awareness of other users necessary to understanding objects that act as part of a larger system. This need demonstrates the strong embodied net of interaction in environments. Found is that context in environments both guide the "rules of use" of a technology while also helping to make the technology understandable. Finally, from their study and work of Dourish [13], we see it is not solely an aid's presence but also the device's orientation that gives environments meaning.

So from these works, we see that the community of practice and cognitive meaning are influenced and guided by the aid technologies within them. It therefore becomes more important that we understand how cognitive meaning will be established *post*-aid introduction, rather than in the current status quo if we are to design more effective cognitive aid systems.

In defining what these design considerations involve, [28] identified three modes of embodied interaction contributing to an object's meaning within an environment of action. *Intrinsic* interaction conveys meaning by how its physical appearance affords certain modes of use [23], and requires no previous knowledge of the object. *Iconic* finds meaning in metaphorical associations we make with objects. People tend to interact with objects as if they were the known objects they depict. For example, an image of the Virgin Mary in tree bark has given the tree deeper meaning to many thereby producing acts of reverence similar to those they have for the actual Virgin Mary. Finally is *instrumental* interaction, where meaning is conveyed through using the object as an instrument. Rather than finding meaning in the physical, meaning is found through using the object as equipment [20] and in the results it produces. Capitalizing on these modes when designing a cognitive aid can result in a more effective and natural cognitive load reduction.

Findings such as these are core to embodied interaction. By understanding how meaning is understood, established, and shared, designers of embodied aids can develop more useable systems for environments. It is not enough to just design for a space without also realizing the social roles and categories established in these places. Ethnographic studies of communities can assist designers in developing systems that are appropriate to the practices and places of embodiment. Further, through an understanding of how cognitive meaning comes about from the actions in which an embodied aid already exists, designers can understand the issues found with existing aid technologies in these communities. Both realizations are important to the design of more usable embodied cognitive aid systems.

Embodied User Interfaces. So we see that it is impossible to engage in environments absent of social meaning and organization. Closely related to this embodiment of interaction is an impossibility of being part of an environment absent of physical structure. Whereas embodied interaction tends to gain most attention in social computing, embodied user interfaces are more in focus as an addition to tangible computing. This is due to the fact that tangible computing finds embodiment through the physical application of it in the world, whereas social computing recognizes embodiment in systems of meaning [13]. Both however, share a presence in the overarching area of embodiment and developers of embodied cognitive aids must also realize that their physical design also communicates meaning.

So we now step aside from focusing on the social interactions of the environment to explore the physical. Prior, we sought to capture cognitive meaning found in the embodied actions of social organization and interaction with aids. Now, we seek to apply that meaning into the structural design of aiding systems. Physical structure infused with both manipulability and representation while still participating in the environment is termed as the embodied user interface. The goal is to move away from consciously "using" to unconsciously "doing", but not necessarily by making an aid's interface invisible, as Dourish confirms [13]. It is the attempt to take the digital equivalent of an object, traditionally only informationally similar, and naturally embed it to make it also interactionally similar, thus freeing cognitive processes.

For example, providing a pilot in combat with a physical or tactile cognitive aid rather than the traditional approach of augmenting cognition through display-based information. Physical controls that were most important at that moment could be physically highlighted while the rest of the controls slightly dimmed to focus cognitive efforts. Better, these controls could take on manipulability functions that are less control-like and more naturally manipulated. The aid during cognitive overload could allow pilot hand motions to depict exactly how they would like the plane to land, with the palm down and fingers slightly up, and then in real-time act upon these motions. Differently, if a pilot was sensed to be incorrectly fixated due to cognitive overload, physical areas of fixation could become tactfully unpleasant by the aid, maybe through rapid temperature change. This tactful approach could release the cognitive fixation and allow refocusing without displaying more visual alert information to an individual that may already be experiencing visual overload. An implemented, however less aid-oriented form of this manipulability is found in the work of Ishii and his colleagues as part of the Tangible Media Group at the MIT Media Lab [24]. Not only does the embodied approach realize meaning through its physical representation, but as we can see, it allows for a more natural description and imagery at the cognitive level. This natural understanding is because the embodied approach is situated in an environment in which we operate rather than within a computer that is confined to few. It is a move that focuses more on providing an aid to cognition by freeing the mind rather than just an augmentation to cognition focused on providing feedback.

So we defined the embodied user interface by using words like *infused*, *manipulability* and *representation*. For Fishkin et al. [15] these features are expressed as tasks "embodied in a device" containing "coincidence of input and output" that "provide highly specific and familiar affordances for particular types of actions". While we add a need for participatory status, it is interesting how these features relate to the intrinsic, iconic, and instrumental forms of interaction mentioned under embodied interaction. This connection reveals further insights for the design of more effective aids.

When Fishkin et al. [15] mention that a task should be "embodied in a device" they are addressing the metaphorical coupling that exists between an embodied aid and an object of familiarity. This associative meaning is what Williams et al. [28] term as the meaning found in iconic interaction. Dourish [13] notes of this feature, that when metaphorical coupling becomes too tight, the metaphor disappears, leaving us with just a simulated activity. Iconic meaning is replaced with direct physical meaning [15] allowing a different, but still favorable design approach. "Familiar affordances for particular types of actions" is the meaning we described earlier in intrinsic interaction. Through this, we see the properties of the physical aid device demonstrating a form of inherent use whether the object is recognizable or not. This is a powerful way to convey meaning simply through physical design. Alternately stated, "coincidence of input and output" is that embodied aids should allow users to input data, interact, or manipulate the device directly, and that outputted feedback should involve a manipulation of the embodied aid. The inputs and outputs should not require a separate computer or computational device that "programs" the object. This is closely related to our earlier concept of meaning involving instrumental interactions with objects. "Coincidences of input and output" are required for aids to become an instrument or equipment effectively used in an assistive way.

Embodied Cognition. Without the social meaning found in embodied interaction and physical meaning embedded in embodied user interfaces, we only have objects, detached from cognitive processes. Do not mistake this as meaning that cognition has no part in embodiment. For embodied cognition realizes that the actions and interactions of the body are an integral part of guiding the mind. This mind impacts the environment by producing changes that come to guide the mind again.

Holding to cognition as a factor of embodiment is a drastic departure from Cartesian cognitive psychology that detaches the mind's processing from the world's activities [12]. Nevertheless, as shown, there are strong arguments in metaphors, affordances, and interaction, against this traditionalist view. So the embodied view of cognition argues that understanding in the mind results from the interactions that take place between individuals and objects situated in environments where one ventures. In our case these objects would be the cognitive aids.

Not only do the world and our actions facilitate meaning in cognition, but we often use the world to help our cognitive processes. This additional dimension also needs consideration in the design of aids as we move from augmenting cognition to aiding cognition. We saw this in our discussions of embodied interaction as both object presence and orientation gave meaning and understanding. But we also often offload and embody cognitive tasks in the world through using object orientation to help us remember and understand particulars of the environment. One place we see this is through the strategic placement and orientation of objects in the environment. Another is by allowing users to manipulate objects in the world rather than requiring it to be done cognitively [17]. Therefore, understanding the embodied nature of cognition proves critical in the effective design of cognitive aids.

Designing aids for embodied cognition involves consideration however that goes beyond the cognitive processing of immediate events. Reconciling two seemingly disconnected views held of Wilson [30], is that off-line cognition, such as what we find in reminiscing, is a process situated in the realm of the embodied environment. This environment is of the past and since has been “freed” from constraints of time. Off-line methods allow us to rethink choices made within the original environment context by using unnoticeable sensorimotor functions [18]. By sensorimotor functioning we mean that motor functioning simulation occurs while using cognitive processes. In essence, a body of embodied interaction is serving the needs of the mind.

We see this taking place in short term memory when we wish to remember a series of numbers. Through repetition, we noticeably or covertly engage motor functioning to facilitate our working memory [21,29]. Alternately, when in time-pressured situations, we often rely on rote memory and long-term recollections held in embodied representations [16]. In fact, much of our indeliberate semantic memory may result from the implicit and episodic memory of interaction [22]. This suggests that the interaction is necessary to our cognitive system and must be considered in aid design.

So the evidence presented speaks strongly to the embodiment of cognition while also showing that meaning comes from the actions and interaction within environments. We know the latter as embodied interaction. Embodying meaning, affordances, and input/output representations in the design of computational cognitive aids then culminates in the creation of embodied user interfaces. To this end we now present an implementation of this type of interface while considering the cognitive and interactive components of embodiment.

4 An Embodied Aid for Space Renovation

It is generally accepted that an ethnographic study of an environment is in order when designing for workspaces. These studies reveal the communities of practice discussed earlier; a contributor to the successful design of embodied user interfaces. In our prototype, we seek to provide a cognitive aid used during space renovations. More specifically, our ethnographic study focuses on the renovation of the Computer Science floor of McBryde Hall at Virginia Tech in Blacksburg, Virginia. Our embodied user interface is based on a set of assumptions and determined an ordinary set of renovation functions by reviewing renovation details of a number of universities [4,6-11]. For illustrative purposes, we present some of the features and design decisions made.

While our aid can assist in providing exact measurements, our focus group is not architects. The client is positioned as a layperson in the renovation cycle and equipped with a set of embodied operations that cognitively aid their decision making as they brainstorm and decide on the configuration of spaces. In the case of a research area of the university these clients are the involved faculty, staff, and students. As a prototype, we do not provide an aid full of all the assistance that an individual "may" need while renovating a space, but rather focus on functionality a layperson "will" need in renovation projects. While our system converges a set of related activities, as our discussion on ubiquitous, social, and mobile computing suggests, we adhere to the concept of a use-designed cognitive aid rather than an all-purpose aid in renovations. Focusing and limiting the functionality allows our system to be more naturally embodied and cognitively accepted by those it assists.

As a first iteration of the aid, we apply insights found in our discussion of mobile computing by equipping the client with a personal digital assistant (PDA) attached to the user's non-writing arm. On their writing hand is a finger-equipped device that intrinsically affords cognitive offloading through writing on the PDA. This device also serves as sensor input for aid processing, a directional tag reader, and laser pointing system. The reader communicates with the tagged PDA using a short-range wireless connection. Each component is based upon existing technologies [1-3,5,25]. To better aid cognition and embody our system, we desired to give a natural form of selection for space configuration rather than relying on a screen-based augmentation approach. A finger-based device was chosen over touch-based, head-mounted, or arm-mounted devices [14] based upon a series of observational studies. We found that a finger-based device affords orientated focus along with control over selection as people naturally refer to items of interest using a finger. Also, this approach allows for smaller levels of movement in specifying start/end sizing and selection points, a more appropriate socio-cultural gesture [15]. Utilizing affordances found in renovations, the client is equipped to design ethnographically as the system is mobile and tangible while remaining unobtrusive. This situatedness reduces cognitive load and encourages the user to better design for a community of interactions.

Rather than offering a new model for renovation, our aid is embedded in current renovation practices. Of course this assumes that computationally mimicking a traditional form of action is the best approach to the action. We however, only take this as a means to an end in developing better aids. The benefit is that mimicking allows the use of current work practices, thereby lowering the cognitive learning curve and allowing the user to work "through" our system, rather than being distracted from their goals by working "on" our system [20]. The system focuses on aiding major decisions that modify internal space use, alter dimensions or configurations, or provide substantial upgrades. From this, we identify three "ends-based" tasks that individuals perform in these decisions. They are adding objects, removing objects, and modifying existing objects. In order to better understand the areas we cognitively aid and augment, we now highlight aspects of these tasks.

In the renovation space we assume that updatable tags have been placed on infrastructural components. As all objects within the space are addressable, including the structural, electrical, mechanical, plumbing, heating, and cooling systems, this provides input to our aiding system. These tags are part of a current system that updates object tag data when changes affecting the object occur, whether directly or indirectly. Typical tag data includes object location, status, installation date, ontological relationships with other same-type objects (e.g. wall joists), and relationships with associated objects (e.g. wires running through the joists).

Our aid relies on a system that allows users to configure building and lab spaces primarily through the use of the multi-purpose finger-based selection tool. In general, extended pointing at an object initiates interaction and returns information on that object, while also allowing the system to provide suggestive input, if desired. While advantages exist to having the cognitive aid suggest ideal configurations, the system, at least initially, allows the user to retain control. As the user is more aware of the climate and interactive nature of the environment, they remain the designer while the aid audibly provides details and feedback on their ideas when cognitively necessary. This design, as tangible computing suggests, keeps the user part of the environment of renovation while also being sensitive to how auditory item lists can retard usability.

Through awareness of the user's location, the system cognitively aids in role-playing how an object could be added into a particular building or lab space. Users select the type of object they are considering in their present location, and use the finger-based directional pointer up or down to naturally specify ceiling or floor placement. When an object for addition exists in the environment, the user selects the object for placement by first pointing at the object. Objects not selectable for addition but inventoried for installation can also be added. Throughout the role-playing process, the cognitive aid uses both the finger-based pointer sensor and voice recognition to assist in cognitive state and decision-making recognition. Voice recognized commands afford a naturally embedded interface approach, and as it is also used in the renovation device, allow for more natural object selection.

Cognitive aid techniques require correspondence with the existing system's relational tag data. As objects of renovation are added, the system aids client processing by also checking previously role-played objects to determine whether the new object's specifications fit within the accepted objects of renovation. If interference or suboptimal placement is noticed, the aid monitors client cognitive processing and offers location or configuration refinement, when necessary. Also, in modifying or deleting objects from the existing space, this relationship data allows the system, if cognitively necessary, to alert a user of issues before an object is modified or removed. An example would be aiding a client if a cognitive slip had them trying to remove a wall before removing the door contained in it.

What we have illustrated are some of the features and design decisions made in creating our embodied cognitive aid. Other aid features exist, and while this paper focuses on the design aspects of embodied aids, we expect the results of tests using this aid along with further design revisions resulting from understanding the changed community of practice to be forthcoming in a future paper.

5 Conclusion

What we have mentioned are some design decisions gained from embodiment upon which our cognitive aid is based. Other insights however are also apparent from our discussion. Offered is a system that does not focus on making decisions, but helps to facilitate those that can. As such, we seek to aid individuals through understanding the communities of practice, rather than replace them. While our embodied interface extends the resources available to individuals, the scope of an available decision at any one time remains natural and manageable. Decision making requires individuals to be immersed in the communities of practice where embodied interaction takes place. By limiting interaction to items within a sensed distance, we reduce the level of conscious interfacing needed and the complexity of the system. Aspects like these make our system a natural embodied experience.

And so our design is based upon our desire to understand embodiment further and how through it we can move beyond simply augmenting cognition to aiding cognition. Evidenced in our discussion is that embodiment is found in many current threads of HCI research. Through a discussion of embodied interaction, user interfaces, and cognition we see that these areas "feed" each other in both the understanding and exploitation of meaning. We can therefore see that embodying computational objects into environments facilitates the creation of usable aids that are immersed with interaction, meaning, and ultimately more effective cognitive aids.

References

1. Bluetooth.com, The Official Bluetooth Wireless Info Site (2005), `http://www.bluetooth.com/bluetooth/` (Accessed January 20, 2006)
2. The Gadgeteer - Finger Tip Styli Comparison Review (2000), `http://www.the-gadgeteer.com/review/finger_tip_styli_comparison_review` (Accessed January 20, 2006)
3. Palm - Products - Smartphones (2006), `http://www.palm.com/us/products/smartphones/` (Accessed January 20, 2006)
4. renovateSTL.com (2006), `http://www.renovatestl.com/booklet.php` (Accessed January 20, 2006)
5. RFID Reader Module | SkyeModule M1-mini (2004) `http://www.skyetek.com/readers_Mini.html` (Accessed January 20, 2006)
6. University of Colorado at Boulder (2006), `http://www.colorado.edu/` (Accessed January 20, 2006)
7. University of Connecticut. (2006), `http://www.uconn.edu/` (Accessed January 20, 2006)
8. University of Delaware (2006), `http://www.udel.edu/` (Accessed January 20, 2006)

9. University of Illinois at Chicago (UIC) (2006), http://www.uic.edu/ (Accessed January 20, 2006)
10. The University of Texas at Austin (2006), http://www.utexas.edu/ (Accessed January 20, 2006)
11. Washington University in St. Louis (2006), http://www.wustl.edu/ (Accessed January 20, 2006)
12. Descartes, R.: Meditationes. Bobbs-Merrill (1641/1964)
13. Dourish, P.: Where the action is: the foundations of embodied interaction. MIT Press, Cambridge (2001)
14. Fallman, D.: Enabling physical collaboration in industrial settings by designing for embodied interaction. In: Proceedings of the Latin American conference on Human-computer interaction. ACM Press, Rio de Janeiro (2003)
15. Fishkin, K.P., Moran, T.P., Harrison, B.L.: Embodied User Interfaces: Towards Invisible User Interfaces. In: Proceedings of the IFIP TC2/TC13 WG2.7/WG13.4 Seventh Working Conference on Engineering for Human-Computer Interaction, Kluwer, B.V. (1999)
16. Glenberg, A.M.: What memory is for. Behavioral and Brain Sciences 20(1), 1–55
17. Glenberg, A.M., Gutierrez, T., Levin, J.R., Japuntich, S., Kaschak, M.P.: Activity and imagined activity can enhance young children's reading comprehension. Journal of Educational Psychology 96(3), 424–436
18. Grush, R.: The emulation theory of representation: motor control, imagery, and perception. Behavioral and Brain Sciences 27, 377–442
19. Harrison, S., Dourish, P.: Re-place-ing space: the roles of place and space in collaborative systems. In: Proceedings of the 1996 ACM conference on Computer supported cooperative work. ACM Press, Boston (1996)
20. Heidegger, M.: Sein und Zeit. Harper & Row, New York (1927/1962)
21. Kosslyn, S.M.: Image and brain: the resolution of the imagery debate. MIT Press, Cambridge (1994)
22. Menon, V., Boyett-Anderson, J.M., Schatzberg, A.F., Reiss, A.L.: Relating semantic and episodic memory systems. Cognitive Brain Research 13(2), 261–265
23. Norman, D.A.: The Design of Everyday Things. Doubleday, New York (1988)
24. Raffle, H.S., Parkes, A.J., Ishii, H.: Topobo: a constructive assembly system with kinetic memory. In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM Press, Vienna (2004)
25. Tsukadaa, K., Yasumurab, M.: Ubi-Finger: Gesture Input Device for Mobile Use. In: APCHI, pp. 388–400 (2002)
26. Weiser, M., Gold, R., Brown, J.S.: The origins of ubiquitous computing research at PARC in the late 1980s. IBM Syst. J. 38(4), 693–696
27. Wenger, E.: Communities of Practice: Learning, Meaning, and Identity. Cambridge University Press, Cambridge (1999)
28. Williams, A., Kabisch, E., Dourish, P.: From Interaction to Participation: Configuring Space Through Embodied Interaction. In: Beigl, M., Intille, S.S., Rekimoto, J., Tokuda, H. (eds.) UbiComp 2005. LNCS, vol. 3660, pp. 287–304. Springer, Heidelberg (2005)
29. Wilson, M.: The case for sensorimotor coding in working memory. Psychonomic Bulletin & Review 8(1), 44–57
30. Wilson, M.: Six Views of Embodied Cognition. Psychonomic Bulletin & Review 9(4), 625–636

CI Therapy: A Method for Harnessing Neuroplastic Changes to Improve Rehabilitation after Damage to the Brain

L.V. Gauthier and E. Taub

Department of Psychology
University of Alabama at Birmingham
lynnevg@uab.edu, etaub@uab.edu

Abstract. Constraint-Induced Movement (CI) therapy has been successfully implemented for treating motor deficit resulting from a variety of previously intractable neurological conditions such as traumatic brain injury. CI therapy's efficacy can be attributed to two interrelated mechanisms: overcoming "learned nonuse" and neuroplasticity. Voxel-based morphometry (VBM) analyses have demonstrated that CI therapy produces lasting *structural* changes to the human brain. Patients that received full CI therapy demonstrated profuse grey matter increases in sensory and motor areas and hippocampus, whereas those who received only intensive motor practice did not. The magnitude of the observed structural changes was correlated with the extent to which the patient regained use of the impaired arm for daily activities. These findings demonstrate that the two mechanisms believed to underlie improvement from CI therapy, overcoming "learned nonuse" and neuroplasticity, act synergistically. Therefore, a bidirectional approach to treating brain injury, one that targets both brain and behavior, is suggested.

1 Introduction

Traumatic brain injury (TBI) is a leading cause of death and lifelong disability among children and young adults in the United States [1]. The number of young Americans experiencing TBI has significantly increased in recent years as a result of combat in Iraq and Afghanistan. According to the Joint Theater Trauma Registry, compiled by the U.S. Army Institute of Surgical Research, over 22% of the soldiers treated in U.S. military hospitals between 2003 and 2005 sustained injuries to the head, neck and face [2]. Approximately 50% of those injured in blast suffered a TBI. Among Walter Reed admissions of war fighters, 54% sustained moderate-severe brain injury [3,4]. Among civilians as well as military personnel, an estimated 5.3 million men, women and children are living with permanent TBI-related disability in the United States today [5]. These disabilities, including movement impairments, often permanently alter a person's vocational capabilities and have profound effects on their lives. The TBI population consists predominately of young males with a potential for employment. Motor disability, combined with cognitive and behavioral deficits, results in poor

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 792–799, 2009.

post-injury employment outcome [6-8]. According to a 1985 study, the annual economic burden of TBI in the United States was approximately $37.8 billion [9]. With rising healthcare costs since that time, these figures have undoubtedly grown substantially. Effective therapeutic interventions for TBI survivors could therefore dramatically decrease TBI's cost to society by increasing quality of life for TBI survivors and their families as well as reducing its economic burden.

A new class of therapy, termed Constraint-induced Movement therapy (CI therapy), was developed to substantially reduce the incapacitating motor deficit and to greatly increase use of an impaired extremity in the life setting following neurological damage. CI therapy was derived from basic behavioral neuroscience research with primates by Taub and colleagues. When a single forelimb is deafferented in a monkey, the animal does not make use of it in the free situation [10,11], but it can be induced to use the deafferented extremity by restricting movement of the intact limb and operant training procedures (e.g. conditioned response training, shaping progressive improvements in movement) [12-14]. Similar training principals were later successfully applied to humans with chronic hemiparesis resulting from stroke; CI therapy has demonstrated efficacy in multiple studies using between- and within-subject controls, placebo controls, and convergent measures from multiple domains [15-20]. A recent multi-site randomized clinical trial (EXCITE) with subacute patients 3-9 months post-stroke reported positive results [19]. Although originally designed to treat stroke patients, CI therapy has been successfully implemented for treating a variety of previously intractable neurological conditions including: TBI [21,22] multiple sclerosis (MS) [23], cerebral palsy [24], and juvenile hemispherectomy [24] with similar clinical outcomes. Although CI therapy does not restore normal movement ability to the arm, individuals who receive the therapy are often able to regain substantially improved use of the formerly hemiparetic arm for many activities of daily living. Its effectiveness may be attributed to two independent but interrelated processes: overcoming learned nonuse of the extremity and altering the structure and function of the human brain.

After any substantial neurological injury, there is a period of suppressed central nervous system activity and corresponding decrease in motor function. During this period, an operant learning process takes place that involves behaviorally reinforced suppression of attempts to move an impaired limb. As a result, individuals with motor deficit demonstrate greatly reduced movement of the affected extremity despite a gradual spontaneous recovery of at least some of the lost function (i.e. learned nonuse) [15, 25]. Almost 90% of TBI patients with motor deficit evaluated in Taub's laboratory exhibited considerably greater use of one arm than the other. Although the brain damage associated with TBI is typically bilateral, damage to the motor network may be greater in one hemisphere than in the other. We believe that this motor advantage of one arm compared to the other is accentuated because the learned nonuse mechanism is based on the greater reward (i.e., success) produced by use of the more effective arm even though the advantage initially may be small. This further increases its use relative to the less used arm, and may result in contractions of the cortical representation zone for the less used arm and increased cortical space devoted to use of the more frequently used extremity [26-30]. This would make the more used extremity progressively easier to use and it would become increasingly difficult for the patient to use the less frequently used arm. The process may be described as a vicious spiral downward for the less used arm that results in the appearance of relative hemiparesis despite bilateral damage to the brain (we have observed this same phenomenon in patients with progressive MS [23], where the neurological damage is also bilateral).

2 Method: CI Therapy

CI therapy consists of three main elements with demonstrated efficacy for overcoming learned nonuse [17, 31]. One component is intensive training of the more affected arm for three hours per day for ten consecutive weekdays. In some respects, this training is similar to what would be obtained through traditional physical or occupational therapy; however, there are several important distinctions: CI therapy focuses entirely on training the more affected extremity, incorporates “shaping” procedures (a desired movement goal is approached in small steps, by successive approximations and continuous feedback), and the duration and intensity of training is greater than is typically carried out in other more traditional forms of therapy. A second component is prolonged restraint of the less affected upper extremity for a target 90% of waking hours to encourage increased use of the more impaired arm. The third and final component is a “transfer package” of behavioral techniques designed to facilitate transfer of therapeutic gains achieved in the laboratory/clinic to real world activities. The transfer package consists of a behavioral contract, monitoring of life situation more-affected arm use by daily administration of the Motor Activity Log (a structured interview concerning the amount and quality of more affected arm use for 30 activities of daily living carried out in the life situation), and problem solving with a therapist to overcome perceived barriers to using the extremity, among several other elements. The transfer package has been shown to be a critical component of the therapy; if absent from the intervention, improvements in spontaneous real world arm use are reduced approximately threefold [32].

3 Results: Neuroimaging Studies Involving CI Therapy

In addition to overcoming learned nonuse, the efficacy of CI therapy may be attributed to a second, related mechanism: neuroplasticity. Decreases in afferent input, such as reduced movement following insult to the brain, have been associated with contraction of the cortical representation zones of the affected extremity [26-30]. Conversely, CI therapy has been shown to increase the cortical representations of affected upper extremity muscles within ipsilesional primary motor cortex in stroke patients [26, 27, 33]. Furthermore, increased recruitment of motor cortex paralleled improvements in amount and quality of daily arm use [27]. CI therapy has been shown to produce “functional” alterations in the excitability, rate of metabolism, and blood flow in ipsilesional brain areas associated with the more affected arm [34, 35]. Other investigators have demonstrated CI therapy-induced functional reorganization in contralesional areas of the brain [36,37], presumably reflecting reorganization of function in the less affected hemisphere. However, a weakness of these studies is that they were limited to using imaging techniques such as transcranial magnetic stimulation [26, 27, 33-35], positron emission tomography [34], and fMRI [35] which record alterations in excitability, rate of metabolism, or blood flow, all of which can fluctuate on a moment-to-moment basis.

More recently, results from a voxel-based morphometry (VBM) study demonstrated that CI therapy produces lasting structural changes to the human brain in addition to the aforementioned physiological changes in brain function [32]. Chronic stroke patients enrolled in a recent randomized controlled trial of CI therapy were assigned to receive either intensive motor practice only (the first component of CI therapy discussed above) or full CI therapy involving all three components including the transfer package. Longitudinal VBM was performed on structural magnetic resonance imaging (MRI) scans obtained immediately before and after patients received therapy to determine structural changes in grey matter. The group receiving all components of CI therapy exhibited far greater improvement in use of the more affected arm in the life situation than the group that received only intensive motor practice. Increases in grey matter paralleled improvements in spontaneous use of the more impaired arm for activities of daily living. The CI therapy group exhibited profuse increases in grey matter in sensory and motor cortices both contralateral and ipsilateral to the affected arm, as well as in bilateral hippocampi. These changes in grey matter were restricted to cortical areas typically involved in motor control of the arm/hand (and not adjacent areas of motor cortex); they may reflect synaptogenesis [38-43], gliosis [44, 45], angiogenesis [46, 47], and possibly neurogenesis [48-51]. The group that received only intensive motor practice failed to exhibit significant grey matter increases. Furthermore, the magnitude of the observed structural changes was correlated with the extent to which the patient regained use of the impaired arm for daily activities. This study demonstrates that real-world arm use is a critical component driving rehabilitation-induced neuroplasticity.

4 Discussion

Although there are several possible explanations for these data, one hypothesis is that neuroplastic changes are sensitive to the behavioral relevance of motor tasks, such as use of the more affected arm for activities of daily living at home encouraged by the transfer package. A similar phenomenon has been demonstrated by Jenkins, Merzenich and colleagues in the sensory system of monkeys [52]. Their studies showed that repetitive "behaviorally relevant" sensory stimulation resulted in plastic expansion of the cortical representations of stimulated digits whereas equal amounts of sensory stimulation that was not behaviorally relevant did not significantly alter these representation zones (behavioral relevance was provided by requiring the monkey to make an accurate discrimination response to differences in the tactile stimulation to obtain food or liquid reward). Alternatively, motor tasks performed in the home may be more complex than the structured tasks used for motor training in the laboratory and may involve the simultaneous coordination of more muscle groups and therefore produce a greater neuroplastic response [39, 40, 46, 53-55]. Empirical investigation could further elucidate the mechanisms by which an individual's behavior influences brain structure and function.

TBI involves different neuropathology than does stroke. The stretching and shearing associated with brain trauma causes a misalignment in the cytoskeleton followed by accumulation of intracellular structures and swelling that can cause a separation of the axon [56]. It is therefore unclear whether the neuroplastic response of TBI patients to CI therapy differs from that demonstrated within a stroke population. However, TBI patients treated with CI therapy show clinical improvements equivalent to those observed in the stroke population [21, 22]. One might make the assumption that behavioral changes are reflective of changes in brain structure or function. Therefore, it is highly possible that neuroplastic mechanisms are operating in the brains of TBI patients treated with CI therapy that are similar to those shown to occur in patients with stroke.

5 Conclusion

What is currently known regarding structural brain changes after CI therapy has substantial implications for rehabilitation. The aforementioned data demonstrates that the two mechanisms believed to underlie improvement from CI therapy, overcoming learned nonuse and neuroplasticity, act synergistically. Although the brain produces movement, purposeful movement can have an equally profound reciprocal effect on brain structure and can be harnessed for therapeutic effect. The magnitude of structural brain change and therapeutic effect appears to depend on the nature and extent of the change in behavior, however. Fundamental behavioral change involving incorporation of the impaired extremity into activities of daily living was necessary to drive neuroplastic reorganization of brain, whereas intensive movement training alone was insufficient. In summary, brain structure/function and the behavior of the individual appear to be interdependent processes that drive therapeutic improvement following insult to the brain. Therefore, a bidirectional approach to treating brain injury, one that targets both brain and behavior, is suggested.

References

1. Sosin, D., Sniezek, J., Thurman, D.: Incidence of mild and moderate brain injury in the united states. Brain Inj. 10, 47–54 (1996)
2. Okie, S.: Traumatic brain injury in the war zone. N. Engl. J. Med. 352, 2043–2047 (2005)
3. Hurley, D.: Neurologists scrambling to keep up with pace of brain injury in Iraq. Neurology Today 5, 5–8 (2005)
4. Warden, D., Ryan, L., Helmick, K.: War neurotrauma: the defense and veterans brain injury center (DVBIC) experience at Walter Reed Army Medical Center (WRAMC) (abstract). J. Neurotrauma. 22, 1178 (2005)
5. CDC. Traumatic brain injury in the United States: a report to Congress (2003)
6. Yasuda, S., Wehman, P., Targett, P., Cifu, D., West, M.: Return to work for persons with traumatic brain injury. Am. J. Phys. Med. Rehabil. 80, 852–864 (2001)
7. Brooks, D., McKinly, W., Symington, C., Beattie, A., Campsie, L.: Return to work within the first seven years of severe head injury. Brain Inj. 1, 5–19 (1987)
8. Gervasio, A., Kreutzer, J.: Kinship and family members' psychological distress after traumatic brain injury: a large sample study. J. Head Trauma Rehabil. 12, 14–26 (1991)

9. Max, W., MacKenzie, E., Rice, D.: Head injuries: costs and consequences. J. Head Trauma Rehabil. 6, 76–91 (1991)
10. Knapp, H., Taub, E., Berman, A.: Effects of deafferentation on a conditioned avoidance response. Science 128, 842–843 (1958)
11. Mott, F., Sherrington, C.: Experiments upon the influence of sensory nerves upon movement and nutrition of the limbs. Proc. R. Soc. Lond. 57, 481–488 (1895)
12. Knapp, H., Taub, E., Berman, A.: Movements in monkeys with deafferented limbs. Exp. Neurol. 7, 305–315 (1963)
13. Taub, E., Berman, A.: Movement and learning in the absence of sensory feedback. In: Freedman, S. (ed.) The neuropsychology of spatially oriented behavior, pp. 173–192. Dorsey Press, Homewood (1968)
14. Taub, E.: Motor behavior following deafferentation in the developing and motorically mature monkey. In: Herman, R., Grillner, S., Ralston, H., Stein, P., Stuart, D. (eds.) Neural control of locomotion, pp. 675–705. Plenum, New York (1976)
15. Taub, E.: Somatosensory deafferentation research with monkeys: implications for rehabilitation medicine. In: Ince, L.P. (ed.) Behavioral Psychology in Rehabilitation Medicine: Clinical Applications, pp. 371–401. Williams & Wilkins, New York (1980)
16. Taub, E., Miller, N.E., Novack, T.A., Cook III, E.W., Fleming, W.C., Nepomuceno, C.S., Connell, J.S., Crago, J.E.: Technique to improve chronic motor deficit after stroke. Arch Phys. Med. Rehabil. 74, 347–354 (1993)
17. Taub, E., Uswatte, G., King, D.K., Morris, D., Crago, J.E., Chatterjee, A.: A placebo controlled trial of constraint-induced movement therapy for upper extremity after stroke. Stroke 37, 1045–1049 (2006)
18. Taub, E.: Harnessing brain plasticity through behavioral techniques to produce new treatments in neurorehabilitation. Am. Psychol. 59, 692–704 (2004)
19. Wolf, S.L., Winstein, C.J., Miller, J.P., Taub, E., Uswatte, G., Morris, D., Giuliani, C., Light, K.E., Nichols-Larsen, D.: Effect of constraint-induced movement therapy on upper extremity function 3 to 9 months after stroke: the excite randomized clinical trial. JAMA 296, 2095–2104 (2006)
20. Morris, D.M., Crago, J.E., DeLuca, S.C., Pidikiti, R.D., Taub, E.: Constraint-induced (CI) movement therapy for motor recovery after stroke. NeuroRehabilitation 9, 29–43 (1997)
21. Shaw, S.E., Morris, D.M., Uswatte, G., McKay, S., Meythaler, J.M., Taub, E.: Constraint-induced movement therapy for recovery of upper-limb function following traumatic brain injury. J. Rehabil. Res. Dev. 42, 769–778 (2005)
22. Shaw, S.E., Morris, D.M., Uswatte, G., McKay, S., Taub, E.: Adherence to constraint induced movement therapy procedures and outcome for persons with traumatic brain injury. J. Neurol. Phys. Ther. 27, 180 (2003)
23. Mark, V.W., Taub, E., Bashir, K., Uswatte, G., Delgado, A., Bowman, M.H., Bryson, C.C., McKay, S., Cutter, G.R.: Constraint-Induced Movement therapy can improve hemiparetic progressive multiple sclerosis. Preliminary findings. Mult Scler. 14, 992–994 (2008)
24. Taub, E., Griffin, A., Nick, J., Gammons, K., Uswatte, G., Law, C.R.: Pediatric CI therapy for stroke-induced hemiparesis in young children. Dev. Neurorehabil. 10, 3–18 (2007)
25. Taub, E., Uswatte, G., Mark, V.W., Morris, D.M.: The learned nonuse phenomenon: implications for rehabilitation. Eura. Medicophys. 42, 241–256 (2006)
26. Liepert, J., Miltner, W.H., Bauder, H., Sommer, M., Dettmers, C., Taub, E., Weiller, C.: Motor cortex plasticity during Constraint-Induced Movement therapy in stroke patients. Neurosci. Lett. 250, 5–8 (1998)

27. Liepert, J., Bauder, H., Wolfgang, H.R., Miltner, W.H., Taub, E., Weiller, C.: Treatment-induced cortical reorganization after stroke in humans. Stroke 31, 1210–1216 (2000)
28. Merzenich, M.M., Nelson, R.J., Stryker, M.P., Cynader, M.S., Schoppmann, A., Zook, J.M.: Somatosensory cortical map changes following digit amputation in adult monkeys. J. Comp. Neurol. 224, 591–605 (1984)
29. Pons, T.P., Garraghty, P.E., Ommaya, A.K., Kaas, J.H., Taub, E., Mishkin, M.: Massive cortical reorganization after sensory deafferentation in adult macaques. Science 252, 1857–1860 (1991)
30. Flor, H., Elbert, T., Knecht, S., Wienbruch, C., Pantev, C., Birbaumer, N., Larbig, W., Taub, E.: Phantom-limb pain as a perceptual correlate of cortical reorganization following arm amputation. Nature 375, 482 (1995)
31. Morris, D.M., Taub, E., Mark, V.W.: Constraint-induced movement therapy: characterizing the intervention protocol. Eura. Medicophys. 42, 257–268 (2006)
32. Gauthier, L.V., Taub, E., Perkins, C., Ortmann, M., Mark, V.W., Uswatte, G.: Remodeling the brain: plastic structural brain changes produced by different motor therapies after stroke. Stroke 39, 1520–1525 (2008)
33. Boake, C., Noser, E.A., Ro, T., Baraniuk, S., Gaber, M., Johnson, R., Salmeron, E.T., Tran, T.M., Lai, J.M., Taub, E., Moye, L.A., Grotta, J.C., Levin, H.S.: Constraint-induced movement therapy during early stroke rehabilitation. Neurorehabil. Neural Repair 21, 14–24 (2007)
34. Wittenberg, G.F., Chen, R., Ishii, K., Bushara, K.O., Taub, E., Gerber, L.H., Hallett, M., Cohen, L.G.: Constraint-induced therapy in stroke: magnetic-stimulation motor maps and cerebral activation. Neurorehabil Neural Repair 17, 48–57 (2003)
35. Hamzei, F., Liepert, J., Dettmers, C., Weiller, C., Rijntjes, M.: Two different reorganization patterns after rehabilitative therapy: an exploratory study with fMRI and TMS. Neuroimage 31, 710–720 (2006)
36. Kopp, B., Kunkel, A., Mühlnickel, W., Villringer, K., Taub, E., Flor, H.: Plasticity in the motor system related to therapy-induced improvement of movement after stroke. Neuroreport 10, 807–810 (1999)
37. Schaechter, J.D., Kraft, E., Hilliard, T.S., Dijkhuizen, R.M., Benner, T., Finklestein, S., Rosen, B.R., Cramer, S.C.: Motor recovery and cortical reorganization after constraint-induced movement therapy in stroke patients: a preliminary study. Neurorehabil. Neural Repair 16, 326–338 (2002)
38. Jones, T.A., Chu, C.J., Grande, L.A., Gregory, A.D.: Motor skills training enhances lesion-induced structural plasticity in the motor cortex of adult rats. J. Neurosci. 19, 10153–10163 (1999)
39. Kleim, J., Barbay, S., Cooper, N.R., Hogg, T., Reidel, C.N., Remple, M.S., Nudo, R.J.: Motor learning-dependent synaptogenesis is localized to functionally reorganized motor cortex. Neurobiol. Learn and Mem. 77, 63–77 (2002)
40. Chu, C.J., Jones, T.A.: Experience-dependent structural plasticity in cortex heterotopic to focal sensorimotor cortical damage. Exp. Neurol. 166, 403–414 (2000)
41. Jones, T.: Multiple synapse formation in the motor cortex opposite unilateral sensorimotor cortex lesions in adult rats. J. Comp. Neurol. 414, 57–86 (1999)
42. Allred, P., Jones, T.: Unilateral ischemic sensorimotor cortical damage in female rats: forelimb behavioral effects and dendritic structural plasticity in the contralateral homotopic cortex. Exp. Neurol. 190, 433–445 (2004)
43. DeBow, S., Davies, M., Clarke, H., Colbourne, F.: Constraint-induced movement therapy and rehabilitation exercises lessen motor deficits and volume of brain injury after striatal hemorrhagic stroke in rats. Stroke 34, 1021–1026 (2003)

44. Briones, T., Woods, J., Wadowska, M., Rogozinska, M., Nguyen, M.: Astrocytic changes in the hippocampus and functional recovery after cerebral ischemia are facilitated by rehabilitation training. Behav. Brain Res. 171, 17–25 (2006)
45. Kleim, J., Markham, J., Vij, K., Freese, J., Ballard, D., Greenough, W.: Motor learning induces astrocytic hypertrophy in the cerebellar cortex. Behav. Brain Res. 178, 244–249 (2007)
46. Black, J., Isaacs, K., Anderson, B., Alcantara, A., Greenough, W.: Learning causes synaptogenesis, whereas motor activity causes angiogenesis, in cerebellar cortex of adult rats. Proc. Natl. Acad. Sci. USA 87, 5568–5572 (1990)
47. Ekstrand, J., Hellsten, J., Tingström, A.: Environmental enrichment, exercise and corticosterone affect endothelial cell proliferation in adult rat hippocampus and prefrontal cortex. Neurosci. Lett. 442, 203–207 (2008)
48. Magavi, E.A., Leavitt, B.R., Macklis, J.D.: Induction of neurogenesis in the neocortex of adult mice. Nature 405, 951–955 (2000)
49. Kolb, B., Morshead, C., Gonzalez, C., Kim, M., Gregg, C., Shingo, T.: Growth factor-stimulated generation of new cortical tissue and functional recovery after stroke damage to the motor cortex of rats. J. Cereb. Blood Flow Metab. 27, 983–997 (2007)
50. Arvidsson, A., Collin, T., Kirik, D., Kokaia, Z., Lindvall, O.: Neuronal replacement from endogenous precursors in the adult brain after stroke. Nat. Med. 8, 963–970 (2002)
51. Wurm, F., Keiner, S., Kunze, A., Witte, O.W., Redecker, C.: Effects of skilled forelimb training on hippocampal neurogenesis and spatial learning after focal cortical infarcts in the adult rat brain. Stroke 38, 2833–2840 (2007)
52. Jenkins, W.M., Merzenich, M.M., Ochs, M.T., Allard, T., Guic-Robles: Functional reorganization of primary somatosensory cortex in adult owl monkeys after behaviorally controlled tactile stimulation. J. Neurophysiol. 63, 82–104 (1990)
53. Nudo, R., Milliken, G., Jenkins, W., Merzenich, M.: Use-dependent alterations of movement representations in primary motor cortex of adult squirrel monkeys. J. Neurosci. 16, 785–807 (1996)
54. Plautz, E., Milliken, G., Nudo, R.: Effects of repetitive motor training on movement representations in adult squirrel monkeys: Role of use versus learning. Neurobiol. Learn and Mem. 74, 27–55 (2000)
55. Kleim, J., Swain, R., Armstrong, K., Napper, R., Jones, T., Greenough, W.: Selective synaptic plasticity within the cerebellar cortex following complex motor skill learning. Neurobiol. Learn and Mem. 69, 274–289 (1998)
56. Arfanakis, K., Haughton, V., Carew, J., Rogers, B., Dempsey, R., Meyerand, M.: Diffusion tensor MR imaging in diffuse axonal injury. AJNR Am. J. Neuroradiol. 23, 794–802 (2002)

Augmented Cognition Design Approaches for Treating Mild Traumatic Brain Injuries

Kay Stanney, Kelly Hale, and David Jones

Design Interactive, Inc.
{kay,kelly,david}@designinteractive.net

Abstract. Augmented cognition could serve as an innovative rehabilitation approach for mild traumatic brain injuries, where issues with cognition, behavior, and affective responses are monitored in real-time and mitigation strategies are triggered to resolve performance or behavior issues. Such mitigations could guide individuals in addressing the current situation (e.g., performance decrement, undesired behavior, negative affective response), as well as provide rehabilitation support to improve performance and behavior in subsequent situations. This paper focuses on mitigation strategies that are suitable for an augmented cognition rehabilitation setting, with the goal of supporting recovery from suboptimal performance and providing rehabilitation tools in real-time, operational context.

Keywords: Augmented cognition, mild traumatic brain injury, mitigation strategies.

1 Introduction

Traumatic brain injury (TBI) is caused by a sudden trauma that leads to damage to the brain. It is a problem affecting many individuals; with an estimated 5 million Americans contending with the challenges TBI brings to daily life [32]. TBI affects almost every dimension of human functioning, with some common disabilities including problems with cognition, sensory processing, communication, and behavior [22]. Specifically, neurophysiological impairments caused by mild TBI have been characterized in terms of three functional areas: (1) information-handling aspects of behavior (e.g., deficits in sequencing information); (2) control aspects of behavior (e.g., doing things slowly to ensure accuracy); and (3) affective response (e.g., expressing feelings of frustration). Deficits in any one or a combination of these areas could negatively impact an individual's ability to return to work and complete everyday work tasks; particularly with similar efficiency and effectiveness as realized prior to experiencing the injury. Rehabilitation is important in the recovery process, where TBI patients seek to return to the normal activities of their daily lives and reenter the workforce. Current approaches to rehabilitation, both cognitive and occupational, focus on helping patients adapt to limitations by providing environmental modifications that simplify vocational activities and/or learning compensatory strategies for coping with chronic cognitive deficits [27]. There is a clear need to

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 800–809, 2009.

develop innovative, biobehaviorally-plausible, and evidence-based approaches to rehabilitation that have the potential to facilitate recovery and the ability to optimize work capabilities for those with mild to moderate TBI.

An innovative rehabilitation approach for addressing this need is augmented cognition, where cognitive state and human performance can be monitored in real-time to identify points of non-optimal performance and activate real-time mitigations that support the human-system dyad [29]. Such mitigations could guide individuals in addressing the current situation (e.g., performance decrement, undesired behavior, negative affective response), as well as provide rehabilitation support to improve performance and behavior in subsequent situations. This paper focuses on mitigation strategies that are suitable for an augmented cognition rehabilitation setting, with the goal of supporting recovery from suboptimal performance and providing rehabilitation tools in real-time, operational context.

Augmented cognition systems to date have focused on optimizing performance of operators in highly complex, information-rich environments. Thus, mitigation frameworks and techniques developed within this domain have focused on optimizing cognitive state (e.g., attention, workload, executive function) and human performance during multi-modal, multi-tasking situations [10]. Within rehabilitation settings, both mitigation frameworks and specific mitigation techniques will likely differ from those in existence today. Specifically, frameworks may have different event triggers or trigger thresholds that indicate a need for system mitigation, and the mitigations built into these frameworks would differ in that they would support the three functional areas impaired in TBI. Information-handling mitigations may focus on such aspects as supporting memory impairments, reduced attention span, and reduced planning capacity; behavioral control mitigations may focus on supporting such aspects as impaired self-monitoring and poor impulse control; and affective response mitigations may focus on such aspects as stress intolerance, frustration, and motivation [22]. Such interventions are intended to not only help TBIs effectively perform vocational activities, but also enable them to progressively acquire or optimize the requisite skills to perform autonomously in the workforce. The current paper discusses TBI-targeted mitigation techniques that could be used to support each of these functional areas.

2 Information-Handling Mitigation Strategies

Cognitive deficits are the most common dysfunction associated with mild TBI and the most detrimental to restoration of normal functioning [19]. Major cognitive impairments associated with mild TBI are memory, attention, and executive function deficits [34]. Information-handling mitigation strategies could help support issues associated with these cognitive dysfunctions.

2.1 Working Memory Impairments

Working memory is an area of information handling that is especially vulnerable to disruption after TBI [21]. Working memory is a temporary storage system requiring attentional resources to retain information from different sensory modalities until a human operator is able to attend to that information. While mild TBI do not tend to have decreases in working memory capacity, per se, instead they tend to experience difficulty with timing and regulating access to information [20, 21]. In addition, those with mild TBI appear to have difficulty allocating processing resources in response to moderate to high working memory tasks. However, if mild TBI are directed to use conscious and deliberate strategies to effectively allocate attentional resources and manage the rate of information during task performance this helps alleviate performance decrements [6]. It is crucial to provide mild TBI with such memory compensatory strategies so they can realize an independent life of reasonable quality.

For working memory, the mitigation challenge becomes – how to assist those with mild TBI in allocating appropriate working memory resources to incoming information and managing the rate of incoming information? Augmented cognition could be used to assist with this working memory allocation and timing problem. For example, Lockheed Martin Advanced Technology Laboratory used electroencephalogram (EEG) indices, galvanic skin response (GSR), pupilometry, and electrocardiogram (ECG), in an augmented cognition system to ease working memory limitations by activating a **sequencing mitigation strategy** [10] that held less critical information in a cue allowing the operator to deal with only critical information during times of high workload [33]. Such a sequencing mitigation strategy could assist mild TBI in allocating working memory resources to critical task information during moderate to high task workload (see Table 1).

Table 1. Information handling mitigations strategies for supporting mild TBI

Mitigation Objective	Mitigation Strategy	Intervention
Support working memory	Sequencing	Allocating attention to critical task information
	Pacing	Managing the flow of task information
Increase attention span	Modality augmentation and switching	Supporting selective or divided attention during performance of complex, functional tasks
Support planning capacity	Cueing	Increasing the number of contextual cues
	Decluttering	Reducing the amount or complexity of information to be displayed to suppress irrelevant information

In terms of timing, Wilson and Russell [36] used a **pacing mitigation strategy** [10], where the speed of display icons (i.e., simulated UAVs) was dynamically reduced depending on the current cognitive load of the operator, providing more time to successfully complete the task at a cost of longer mission duration. Such a pacing mitigation strategy could assist mild TBI in managing the flow of task information during moderate to high task workload. Once information enters working memory, attentional resources must be directed toward the process of retaining a memory and supporting human performance.

2.2 Reduced Attention Span

Attentional deficits in mild TBI can affect arousal/alertness, selective and divided attention, and energetic components of attention (allocation and speed of processing; [23]). These individuals may experience difficulty concentrating on even simple tasks or dividing their attention among multiple tasks [9]. Sohlberg et al. [30] found evidence supporting the effectiveness of direct attention training after TBI on complex, functional tasks requiring selective or divided attention. Training with different stimulus modalities has shown particular promise in attention training for those with mild TBI [5]; [7]. This type of intervention could readily be supported by an augmented cognition system. Specifically, an augmented cognition system could monitor attention deficits in task demands (e.g., difficulty dividing attention during multi-tasking). Mitigations strategies could then be invoked that effectively redirect attention through the use of varied stimulus modalities. Such an approach is known as **modality augmentation and switching** [10], which can occur in two forms: (1) Modality redundancy refers to presenting the same information in multiple modalities, and/or providing complimentary information in a second modality, and (2) Modality switching replaces one sensory modality with another in order to optimize distribution of processing load. Jones (2005) explored the utility of modality augmentation in a simulated unmanned aerial vehicle (UAV) task (note, the system was not a closed-loop). The integration of audio cues (as compared to a visual display alone) led to increased detection rates of critical mission events and increases in speed of detection, while reducing the perceived workload associated with the task. Furthermore, these gains in turn led to increases in accuracy while performing tasks after attention was directed to them. This and other such successes indicate that attentional deficits are an appropriate and promising focus area for augmented cognition rehabilitation efforts.

2.3 Reduced Planning Capacity

Those with mild TBI can experience difficulty in planning and executing tasks, especially multiple coordinated tasks [9]. Such tasks require conceptual and executive functioning abilities, which are diminished in those with mild TBI. These deficits can make it difficult to devise and follow a sequence of job tasks to completion, especially when such tasks require flexibility and creative problem solving. Executive function in human information handling is responsible for directing attention, suppressing irrelevant information, coordinating cognitive processes during multi-tasking, and handling novel situations (i.e., those with contextual ambiguity; [11]). An augmented cognition system could support the executive function of those with mild TBI by identifying inattention during task activity and reducing contextual ambiguity, and thus effortful executive function processing, by increasing the number of contextual cues (i.e, cueing; [10]) associated with a target task thereby assisting operators in quickly reinstating context during multi-tasking. For example, Boeing [2] used an adaptation strategy involving **cued retrieval** and **decluttering**, which facilitated the chunking of related information and suppression of irrelevant information in a UAV display during high levels of decision-making. Amplification of the executive function was also used in the Cognitive Cockpit, where a task interface manager adaptively provided cueing information according to the situational context and pilot state [31]. Cueing and decluttering mitigation strategies could assist mild TBI in planning and executing tasks during moderate to high task workload.

3 Behavorial Control Mitigation Strategies

Lowered self-monitoring ability and impulsivity are nettlesome behavioral problems associated with mild TBI [19]. Behavioral control mitigation strategies could support such dysfunctions.

3.1 Self-monitoring of Behavior

Self-monitoring of behavior is commonly disrupted in those with mild TBI [14]. Self-monitoring is an ongoing self-assessment involving comparison of actual versus expected performance that occurs when an individual is engaged in a task activity. Through self-monitoring, an individual generates internal feedback that can be used to self-regulate their performance. In mild TBI, self-monitoring accuracy does not tend to generalize across diverse task types and thus interventions often have to be task specific. Increases in self-monitoring ability have been realized by showing an individual a replay of their task performance [3], while engaging them in metacognitive strategies (reflection, prediction of difficulty of next steps, comparing actual with expected performance; [19]), as well as by providing verbal and visual performance feedback – such as auditory prompts during initial learning - and praise [26]. These types of interventions could be readily incorporated into an augmented cognition system. For example, Bose, van Doesburg, van Maanen, and Treur [4] used cognitive models to augment metacognition in an augmented cognition system for an air traffic control task requiring visual attention. Specifically, eye gaze was used to detect if an operator was paying attention (i.e., actual performance) to the most threatening radar tracks (i.e., expected performance) and where deviations were found tasks were allocated to automation. Similarly, Hudlicka [12] developed a cognitive architecture module that performs metacognitive functions involved in monitoring and control of cognition. The module modeled the assessment and projection of a current situation onto expectations and prediction of possible future states. It also directed actions when the actual situation was not in line with expectations through **metacognitive prompting and feedback**, including prompting to re-scan the environment, obtain contextual and task feedback, and assess the situation. Such a module could be incorporated into a rehabilitation system to support the self-monitoring behavior of those with mild TBI (see Table 2).

Table 2. Behavioral control mitigations strategies for supporting mild TBI

Mitigation Objective	Mitigation Strategy	Intervention
Support self-monitoring of behavior	Metacognitive prompting and feedback	Encourage ongoing self-assessment involving comparison of actual versus expected performance
Support impulse control	Multimodal warnings	Release stimulus boundedness and foster holistic situation assessment to avert impulsivity

3.2 Poor Impulse Control

Individuals who have suffered mild TBI often an issue with impulsivity, which may manifest itself through snap decisions and poor judgment [19]. This poor impulse control is often due to a tendency for those with mild TBI to become overly preoccupied with the most salient cue(s) in an environment, without regard to a holistic assessment of a situation (i.e., stimulus boundedness). This tendency to respond to fragmentary, immediate experiences can also lead to a difficulty in shifting focus from one task to another. For such poor impulse control, the mitigation challenge thus becomes – how to assist those with mild TBI in overcoming stimulus boundedness and redirect their attention to a more holistic situation assessment. **Multimodal warnings** have been effectively used in augmented cognition systems as mitigations that redirect an operator's attention from the current task to a more critical task and thus may prove effective in resolving stimulus boundedness. For example, Barker et al. [2] used an auditory warning that directed an operator's attention to a Tactical Situation Display to search for time-critical targets. Mathiak et al. [18] suggest that crossmodal effects can be used to enhance perception of such warning cues even under demanding task situations. Thus, multimodal warnings could be used in augmented cognition rehabilitation systems to try to release an individual with mild TBI from stimulus boundedness and foster a more holistic situation assessment in an effort to avert impulsivity (see Table 2).

4 Affective Response Mitigation Strategies

Those with mild TBI often experience affective instability, such as lower stress tolerance, reduced frustration tolerance, and lowered motivation [8]. Affective response mitigation strategies could support such issues.

4.1 Reduced Stress Tolerance

Those with mild TBI have reduced tolerance to stress [8]. Cognitive behavioral therapy (CBT) is a common treatment for such stress disorders. CBT involves stress inoculation (i.e., education about reactions to stress, imaginal exposure to stress-inducing memories, cognitive restructuring) and graded in vivo exposure to stressful re-experiences [15]. Such CBT techniques could be incorporated into augmented cognition systems that provide adaptive therapy based on the state of the operator. Such a system could allow an individual to initially practice a task without stressors. Then, stress activators could be introduced to allow the individual to develop coping strategies and review their performance, the latter of which has been shown to significantly improve the effectiveness of stress inoculation training [28]. To effectively support this approach in an adaptive rehabilitation system, stress levels would be monitored in real-time and adaptive rehabilitation would be triggered using a scaffolding approach, ensuring that individuals are performing under low-stress conditions initially and increasing stress as level of mastery increases (see Table 3).

Table 3. Affective response mitigations strategies for supporting mild TBI

Mitigation Objective	Mitigation Strategy	Intervention
Stress inoculation	Phased stress inoculation	Phased stress exposure, adjusting scenarios from initially low-stress conditions, ultimately challenging with high-stress conditions
Frustration tolerance	Frustration release strategies	Simplify tasks, slow pace, provide a calming environment (e.g., music), encourage breaks in activity
Motivation enhancement	Motivational techniques	Verbal motivation and goal visualization techniques

4.2 Diminished Frustration Tolerance

Diminished frustration tolerance is a common ailment for those with mild TBI [9]. Such individuals readily feel overwhelmed by their task circumstances. Frustration can be mediated by simplifying tasks, slowing the pace, providing a calming environment (e.g., music), and encouraging breaks in activity when frustration is evident [5, 7. 15]. Further, Liebman [16] notes that when an individual experiences unnecessary or excessive frustration he tends to become psychologically fixed at the point of frustration. Augmented cognition rehabilitation systems should thus be designed to monitor for and release such fixation through the use of breaks or other such techniques. Augmented cognition systems could thus incorporate mitigation strategies that serve as frustration release techniques that support rehabilitation (see Table 3).

4.3 Motivation

Apathy is a common ailment associated with mild TBI, which can manifest via disinterest in daily task activities [25]. Such increased disinterest and decreased motivation can be persistent, lingering for years after a brain injury [17]. Verbal motivation and goal visualizing techniques have shown promise in enhancing the motivation of those with mild TBI [13]. Such motivation techniques (e.g., positive feedback, knowledge of goals/ performance) could be incorporated into mitigations strategies for an augmented cognition rehabilitation system. For example, Qu, Wang, and Johnson [24] used eye gaze to characterize a learner's attention and then infer motivation factors of the learner in an interactive learning environment. Further, facial expressions could be monitored in real-time and used to characterize motivation. For example, Adolphs [1] characterized emotions as discrete entities and temporary intersections within a continuum of states defined in terms of motivation and alertness. Once an individual's motivation state was determined, then verbal motivation and goal visualization techniques could be incorporated into mitigation strategies that adapt rehabilitation in real-time to foster greater motivation (see Table 3).

5 Mitigation Framework

In general, the mitigations strategies described above would be designed as part of a rehabilitation system, which provides compensatory tools and techniques that adapt to and are used by the individual to allow functioning in spite of disabilities and restoration through repetitive exercises and activities designed to restore or improve damaged abilities. Further, as opposed to previously developed mitigation frameworks that adopted a step-wise approach, where minimally invasive mitigation techniques were employed prior to more intrusive techniques - with the goal of ensuring mitigation benefits outweighed potential adverse effects [10], a rehabilitation mitigation framework would turn this step-wise approach on its head, providing more intrusive mitigations initially to correct suboptimal performance or behavior, and subtly reducing the amount of assistance over time as cognitive state and performance indicators show improved capability to handle increased difficulty of tasks. Such a rehabilitation mitigation approach is similar to past "training wheels" paradigms, where initially a high-level of support is provided (i.e., automated sequencing of steps, slowed information presentation rate), then it is scaled back (i.e., manual sequencing of steps, slowed information presentation rates), and then scaled back again (i.e., manual sequencing of steps, restored information presentation rates). While one could base this progression on overt behavioral responses, this would be an informed "guess" based on performance outcomes. The use of augmented cognition supports precision rehabilitation; i.e., presenting / removing mitigations at precisely the right moment – when the individual demonstrates limited or increasing cognitive capacity, impulsivity, stress, etc. at the neuropsychological level. This would allow for targeting of requisite neuroplasticity in the brain that drives rehabilitation.

References

1. Adolphs, R.: Recognizing emotion from facial expressions: Psychological and neurological mechanisms. Behavioral and Cognitive Neuroscience Reviews 1, 21–62 (2002)
2. Barker, R.A., Edwards, R.E., O'Neill, K.R., Tollar, J.R.: DARPA Improving Warfighter Information Intake Under Stress – Augmented Cognition Concept Validation Experiment Analysis Report for the Boeing Team (NBCH030031). DARPA, Arlington (2004)
3. Beer, J.S., John, O.P., Scabini, D., Knight, R.T.: Orbitofrontal cortex and social behavior: integrating self-monitoring and emotion-cognition interactions. Journal of Cognitive Neuroscience 18(6), 871–879 (2006)
4. Bose, T., van Doesburg, W., van Maanen, P.-P., Treur, J.: Augmented metacognition addressing dynamic allocation of tasks requiring visual attention. In: Schmorrow, D.D., Reeves, L.M. (eds.) Foundations of Augmented Cognition, Proceedings of 3rd Augmented Cognition International, Beijing China, July 22-27, pp. 166–175 (2007)
5. Cappa, S.F., Benke, T., Clarke, S., Rossi, B., Stemmer, B., van Heugten, C.M.: EFNS guidelines on cognitive rehabilitation: report of an EFNS Task Force. European Journal of Neurology 12, 665–680 (2005)
6. Cicerone, K.D.: Remediation of 'working attention' in mild traumatic brain injury. Brain Injury 16(3), 185–195 (2002)
7. Cicerone, K.D., Dahlberg, C., Kalmar, K., Langenbahn, D.M., Malec, J.F., Bergquist, T.F., Felicetti, T., Giacino, J.T., Harley, J.P., Harrington, D.E., Herzog, J., Kneipp, S., Laatsch, L., Morse, P.A.: Evidence based cognitive rehabilitation: recommendations for clinical practice. Archives of Physical Medicine and Rehabilitation 81(12), 1596–1615 (2000)
8. Crowe, S.F.: The behavioural and emotional complications of traumatic brain injury. CRC Press, Boca Raton (2008)

9. Fabiano, R.J., Daugherty, J.: Rehabilitation considerations following mild traumatic brain injury. Journal of Rehabilitation 64(4), 9–14 (1998)
10. Fuchs, S., Hale, K.S., Stanney, K.M., Juhnke, J., Schmorrow, D.: Enhancing mitigation in augmented cognition. Journal of Cognitive Engineering and Decision Making 1(3), 309–326 (2007)
11. Grafman, J.: Executive functions. In: Parasuraman, R., Rizzo, M. (eds.) Neuroergonomics: The Brain at Work, pp. 159–177. Oxford University Press, Oxford (2007)
12. Hudlicka, E.: Modeling interaction between metacognition and emotion in a cognitive architecture. In: Proceedings of the AAAI Spring Symposium on Metacognition in Computation, AAAI Technical Report SS-05-04, pp. 55–61. AAAI Press, Menlo Park (2005)
13. Kellera, M., Hiltbrunnerb, B., Dilla, C., Kesselring, J.: Reversible neuropsychological deficits after mild traumatic brain injury. Journal of Neurology, Neurosurgery, and Psychiatry 68, 761–764 (2000)
14. Kennedy, M.R.T., Coelho, C.: Self-regulation after traumatic brain injury: a framework for intervention of memory and problem solving. In: Turkstra, L.S. (ed.) Seminars in Speech and Language, vol. 26(4), pp. 242–255 (2005)
15. Kornør, H., Winje, D., Ekeberg, O., Weisæth, L., Kirkehei, I., Johansen, K., Steiro, A.: Early trauma-focused cognitive-behavioural therapy to prevent chronic post-traumatic stress disorder and related symptoms: A systematic review and meta-analysis. BMC Psychiatry 8, 81 (2008)
16. Liebman, S.: Stress situations. Lippincott, Philadelphia (1955)
17. Marin, R.S., Wilkosz, P.A.: Disorders of diminished motivation. Journal of Head Trauma Rehabilitation 20, 377–388 (2005)
18. Mathiak, K., Hertrich, I., Świrszcz, K., Zvyagintzsev, M., Lutzenberger, W., Ackermann, H.: Preattentive crossmodal information processing under distraction and fatigue. In: Schmorrow, D.D. (ed.) Foundations of Augmented Cognition, pp. 1055–1064. Lawrence Erlbaum Associates, Mahwah (2005)
19. McAllister, T.W.: Neurobehavioral sequelae of traumatic brain injury: evaluation and management. World Psychiatry 7(1), 3–10 (2008)
20. McAllister, T.W., Saykin, A.J., Flashman, L.A., Sparling, M.B., Johnson, S.C., Guerin, S.J., Mamourian, A.C., Weaver, J.B., Yanofsky, N.: Brain activation during working memory 1 month after mild traumatic brain injury: A functional MRI study. Neurology 53(6), 1300–1308 (1999)
21. McAllister, T.W., Sparling, M.B., Flashman, L.A., Guerin, S.J., Mamourian, A.C., Saykin, A.J.: Differential working memory load effects after mild traumatic brain injury. Neuroimage 14(5), 1004–1012 (2001)
22. National Institute of Neurological Disorders and Stroke. Traumatic brain injury: hope through research. National Institutes of Health, Bethesda. NIH Pub No.: 02-158 (2002)
23. Niemann, H., Ruff, R.M., Kramer, J.H.: An attempt towards differentiating attentional deficits in traumatic brain injury. Neuropsychology Review 6(1), 11–46 (1996)
24. Qu, L., Wang, N., Johnson, W.L.: Using learner focus of attention to detect learner motivation factors. In: Ardissono, L., Brna, P., Mitrović, A. (eds.) UM 2005. LNCS, vol. 3538, pp. 70–73. Springer, Heidelberg (2005)
25. Rao, V., Spiro, J.R., Schretlen, J.D., Cascella, N.G.: Apathy syndrome after traumatic brain injury compared with deficits in schizophrenia. Psychosomatics 48, 217–222 (2007)
26. Rebann, M., Hannon, R.: Treatment of unawareness of memory deficits in adults with brain injury: three case studies. Rehabilitation Psychology 40, 279–287 (1995)

27. Rohling, M.L., Faust, M.E., Beverly, B., Demakis, G.: Effectiveness of cognitive rehabilitation following acquired brain injury: A meta-analytic re-examination of Cicerone et al.'s (2000, 2005) systematic reviews. Neuropsychology 23(1), 20–39 (2009)
28. Saunders, T., Driskell, J., Johnston, J., Salas, E.: The Effects of Stress Inoculation Training on Anxiety and Performance. J. of Occupational Health Psych. 1(2), 170–186 (1996)
29. Schmorrow, D.D., Stanney, K.M.: Augmented Cognition: A Practitioner's Guide. Human Factors and Ergonomics Society, Santa Monica (2008)
30. Sohlberg, M.M., Avery, J., Kennedy, M., Ylvisaker, M., et al.: Practice guidelines for direct attention training. Journal of Medical Speech-Language Pathology 11, 19–39 (2003)
31. Taylor, R., Bonner, M., Dickson, B., Howells, H., Miller, C., Milton, N., Pleydell-Pearce, K., Shadbolt, N., Tennison, J., Whitecross, S.: Cognitive cockpit engineering: Coupling functional state assessment, task knowledge management, and decision support for context-sensitive aiding. In: McNeese&, M., Viculich, M. (eds.) Cognitive systems engineering in military aviation environments: Avoiding cogminutia fragmentosa!, pp. 253–314. Human Systems Information Analysis Center Press, Wright Patterson AFB, Ohio (2002)
32. Thurman, D., Alverson, C., Dunn, K., Guerrero, J., Sniezek, J.: Traumatic brain injury in the United States: a public health perspective. Journal of Head Trauma Rehabilitation 14(6), 602–615 (1999)
33. Tremoulet, P., Barton, P., Craven, P., Corrado, C., Mayer, G., Stibler, K., et al.: DARPA Improving Warfighter Information Intake Under Stress – Augmented Cognition Phase 3 Concept Validation Experiment (CVE) Analysis Report for the Lockheed-Martin ATL Team (Prepared under contract No. NBCH030032). DARPA/IPTO, Arlington (2005)
34. Trudeau, D.L., Anderson, J., Hansen, L.M., Shagalov, D.N., Schmoller, J., Nugent, S., Barton, S.: Findings of mild traumatic brain injury in combat veterans with PTSD and a history of blast concussion. The Journal of Neuropsychiatry and Clinical Neurosciences 10, 308–313 (1998)
35. Vining Radomski, M.: Traumatic brain injury. In: Vining Radomski, M., Trombly Latham, C.A. (eds.) Occupational Therapy for Physical Dysfunction, 6th edn., pp. 1042–1078. Lippincott Williams & Wilkins, Baltimore (2007)
36. Wilson, G.F., Russell, C.A.: Psychophysiologically versus task determined adaptive aiding accomplishment. In: Schmorrow, D.D., Stanney, K.M., Reeves, L.M. (eds.) Foundations of augmented cognition, 2nd edn., pp. 201–207. Strategic Analysis, Inc., Arlington (2006)

Brain Processes and Neurofeedback for Performance Enhancement of Precision Motor Behavior

B. Hatfield, A. Haufler, and J. Contreras-Vidal

Department of Kinesiology
University of Maryland, College Park, MD 20742
{bhatfiel,haufler,pepeum}@umd.edu

Abstract. Based on a number of empirical investigations of cerebral cortical dynamics during precision aiming tasks (i.e. marksmanship) employing electroencephalography (EEG) refinement of cortical activity and attenuation of non-essential cortico-cortical communication with the motor planning regions of the brain results in superior performance. Employment of EEG neurofeedback during the aiming period of target shooting designed to reduce cortical activation resulted in improved performance in skilled marksmen. Such an effect implies that refinement of cortical activity is causally related to performance. Recently, we examined cerebral cortical dynamics during the stress of competitive target shooting and observed increased activation and cortico-cortical communication between non-motor and motor regions relative to a practice-alone condition. As predicted, this finding was associated with degradation of shooting performance. These findings imply that neurofeedback targeted to brain regions related to emotional responding may preserve the cortical dynamics associated with superior performance resulting in improved accuracy of precision aiming performance.

Keywords: electroencephalography (EEG), psychomotor performance, cognitive neuroscience, stress, kinematics.

1 Introduction

1.1 Cognitive Neuroscience of Skilled Motor Performance – Cortical Dynamics

Based on a number of empirical investigations of cerebral cortical dynamics during precision aiming tasks (i.e. marksmanship) employing electroencephalography (EEG) we have formulated a model of psychomotor efficiency. In essence, this model posits that refinement of cortical activity and attenuation of nonessential cortico-cortical communication with the motor planning regions of the brain results in superior performance. This is likely due to "simplification" of central processes that emerges as economy and consistency of limb actions. More specifically, we have noted a remarkable relationship between left temporal and parietal activity during the aiming period such that quiescence or "relaxation" in this region is related to higher-quality performance.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 810–817, 2009.

1.2 Expert vs. Novice EEG Contrasts

This principle was clearly supported by Haufler et al. [1]. Specifically, novice and expert marksmen were subjected to a target-shooting task as well as comparative verbal (i.e., word recognition) and spatial tasks (i.e., dot localization), with which the groups were similar in terms of experience, while recording EEG. The verbal and spatial tasks were also performed in the shooting stance posture. As shown in the three panels below, lower cortical activation levels in the cerebral cortex were observed in the experts during the aiming period of shooting, as measured by gamma 40-Hz power, while no differences were revealed during the comparative tasks. The left-sided and middle panels of Figure 1 represent comparative log-transformed gamma power (36-44 Hz), which is also positively related to cortical activation, from the averaged homologous frontal (*Panel A*) and the averaged homologous temporal regions (*Panel B*). The group differences in activation associated with the frontal region also suggest that the experts were less reliant on effortful executive processing (i.e., planning and coordinating processes) as compared to their novice counterparts. Experts also revealed significantly lower beta power, a spectral band that is positively related to cortical activation, during shooting while, again, no differences were noted during the verbal and spatial tasks. *Panel C* shows that the experts exhibited higher levels of alpha power (8-12 Hz) at site T3 as well as lower levels of beta and gamma power. Collectively, the results clearly show task-specific relaxation in the cortex of the expert marksman.

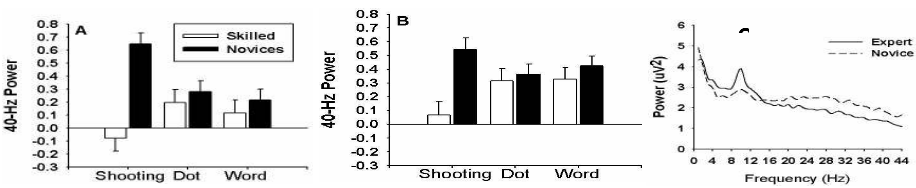


Fig. 1. Expert novice contrasts of EEG spectral power: (A) bilateral frontal gamma power (F3, F4), (B) bilateral temporal gamma power (T3, T4), (C) spectral power at T3 during shooting

1.3 Stress and Cortical Dynamics

Although these expert-novice contrasts and training studies did not involve direct manipulations of psychological stress, they do provide support for the notion that the individual who is highly skilled and focused shows suppression of task-irrelevant associative activity and concomitant activation of task-relevant processes [2]. The notion of mental economy is a fundamental building block of the neurobiological model of superior performance and the antithesis to the state observed during anxiety. The novice or one who is marked by a lack of confidence and much uncertainty of their performance would theoretically exhibit hyperactivity of numerous cortical processes. In this regard Kerick, Hatfield and Allender [3] recently observed a positive relationship between cortical activity (i.e., alpha suppression) and cognitive load in a study of U. S. Marines executing a target-shooting task.

Cortical coherence and shooting performance. Additional insight can be attained into the neurobiology of the skilled performance state by examination of functional interconnectivity or cortico-cortical communication between specified topographical regions of the brain. Such "networking" activity can be quantified by deriving coherence estimates between selected pairs of electrodes or recording sites [4]. In a recent study Deeny et al. [5] assessed inter-electrode coherence between motor planning (Fz) and association areas regions of the brain during skilled marksmanship by monitoring EEG at sites F3, F4, T3, T4, P3, Pz, P4 as well as the motor cortex (C3, Cz, C4) and visual areas (O1 and O2). Coherence was assessed during a 4-second aiming period just prior to trigger pull in two groups of participants who differed in competitive performance history (experts and non-experts). The two groups were equally experienced (approximately 18 years of practice), but the "experts' consistently scored higher under the stress of competition. Figure 2 illustrates the left hemisphere Fz-F3, Fz-C3, Fz-P3, Fz-T3, and Fz-O1 coherence estimates contrasted between the two groups. A significant difference between the groups was detected for the Fz-T3 alpha band coherence, as experts revealed significantly lower values, although no other differences were observed for either the left or right hemisphere. The general lack of group differences in cortical networking seems reasonable as both were similarly experienced with the task and challenged in a similar manner. The Fz-T3 results suggest that experts limited communication between verbal-analytic and motor control processing, thereby simplifying motor planning and performing in a more accurate and consistent manner.

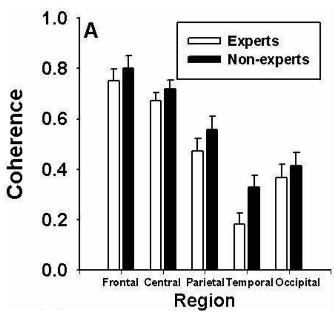


Fig. 2. The left hemisphere Fz-F3, Fz-C3, Fz-P3, Fz-T3, and Fz-O1 coherence estimates contrasted between the two groups is illustrated

Collectively, the results of these studies suggest that superior performance is marked by mental economy, particularly of analytical associative processes, and that pruning of excessive cortico-cortical communication between such processes and motor regions underlies enhancement and consistency of psychomotor (shooting) performance and raise the prediction that a reversal of these patterns would be observed with the imposition of stress.

1.4 Affective Neuroscience - Brain Processes during Emotion

Bear, Conners, and Paradiso [6] recently summarized the neural structures involved in a system or circuit, which mediates the psychological and physiological response to stress. Generally, the stress response is orchestrated by the limbic system but the central components of this functional circuit are the amygdalae, small almond-shaped structures located bilaterally and anterior to the hippocampi on the inferior and medial aspect of the temporal lobes [6]. Multiple sensory pathways converge in the basal lateral nuclei of the amygdalae so that environmental events are immediately processed. Depending on the valence of the stimuli, the lateral nuclei then communicate with the central nucleus in each amygdala and subsequent connections travel to critical forebrain, brainstem, autonomic, and endocrine structures that mediate the expression of emotion. Specifically, there are interconnections from the central nuclei to the (1) hypothalamus, which results in sympathetic arousal and stimulation of stress hormones via the hypothalamic-pituitary-adrenocortical (HPA) axis, (2) the periaqueductal grey, which results in motor responses, and (3) the cingulate cortex, which results in additional cortico-cortical communication with neocortical association regions such as the temporo-parietal regions.

In this manner orchestrated sequelae occur in response to a stressful environment, which, collectively, can change the performer's mental and physical state in a profound manner. For example, heart rate and cortisol levels rise, as does muscle tension, and the soldier may concomitantly experience excessive self-talk and "too much thinking" such that their attention is compromised and the execution of normally automated psychomotor skills such as marksmanship become explicitly managed - timing and coordination are then altered and likely reduced in quality while attention shrinks.

In light of the mental and physical changes that accrue, the activation of the amygdalae serves as a pivotal event in the manifestation of stress and the control of activity in the amygdalae would exact a powerful influence on the performer's mental and physical state. Beyond the structures and processes outlined by Bear et al. [6] a critical component of the neurobiology of fear Importantly, the anterior cortical regions have extensive anatomical connections with several subcortical limbic structures implicated in emotional behavior, particularly the amygdala**.** Davidson and colleagues [7,8] have generated a significant body of literature that clearly shows a positive association between left frontal activation and positive affect while relative right activation is associated with negative affect [7]**.** Although the lateralization of frontal activation is robustly related to the valence of emotion as described above, recent evidence points to a more fundamental association such that left frontal activation mediates approach-oriented behavior while right frontal activation is associated with avoidance or withdrawal-oriented behavior [8,9]. For example, left frontal activation is manifest during hostile behavior, which is certainly not a positive affective state, but most definitely involves approach toward an intended target. Whether positive in nature, approach-oriented, or a combination of the two dimensions, it would appear that such a neurobiological state would be highly adaptive for the individual who must control his/her arousal level while actively engaged with challenging tasks while under great mental stress. Therefore, cortical activation in the frontal region provides an opportune target for neurofeedback training to enable a heightened level of executive control over emotional response and task engagement during challenge.

1.5 Model of Stress-Induced Cortical Dynamics

Figure 3 below provides a model of the processes and outcomes underlying stress reactivity and integrates affective and cognitive activity with psychomotor performance. A central tenet is that lack of executive control over subcortical processes would result in heightened emotional influence (limbic structures) that, in turn, disrupt higher cortical association processes that resulting in alterations in the activation of the motor loop – the fronto-basal ganglia structures that initiate and execute movement. Such disregulation interferes with attention and the motor loop connections (i.e., basal ganglia) to the motor cortex that largely control corticospinal outflow and the resultant quality of motor unit activation [10]. Excessive networking in the cortex may result in undesirable alterations in information processing as well as inconsistency of motor performance. In this manner the motor cortex becomes "busy" with excessive input from limbic processes via increased neocortical activity in the left hemisphere then inconsistent motor behavior would likely result [5].

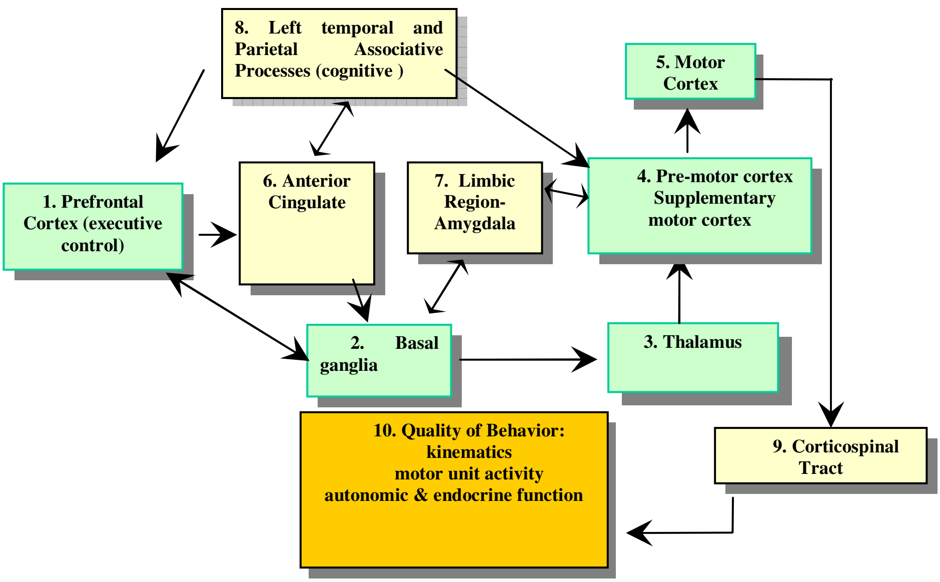


Fig. 3. Model of stress-induced cortical dynamics

Refinement or economy of cortical activation would more likely result in enhanced attention and smooth, fluid, graceful, and efficient movement. Any reduction of associative networking with motor control processes would also help to reduce the complexity of motor planning and should result in greater consistency of performance.

According to this model individuals under high stress will exhibit reductions in prefrontal asymmetry (box 1) compared to a low-stress condition implying a lack of executive control over the fronto-meso-limbic circuit. Consequently, participants will experience heightened activation of the limbic region (amygdala) (box 7). The resultant emotional reactivity, in turn, will result in EEG alpha desynchrony particularly in the left temporal (T3) and parietal (P3) regions (box 8) along with increased cortico-cortical communication between these regions and the motor planning centers (box 4). Such disregulation of the cerebral cortex will be expressed as inconsistent input to the motor loop (boxes 2 - 5) resulting in inconsistent corticospinal output and shooting performance (motor unit activity – trigger pull – boxes 9 and 10). It is well established that attention capacity shrinks with arousal and, consistent with this notion, the excessive cortico-cortical networking during heightened stress, as proposed here, would compromise information processing. In addition, cardiovascular activity (vagal tone) will be inversely related to the activity in the CNS such that vagal tone will be reduced in the high-stress condition. Cortisol levels will rise. The magnitude of change specified in the model will be related to degradation in shooting performance (i.e., slower and inaccurate).

2 Methods

2.1 Participants

Members of the Reserve Officers Training Corps (ROTC) participated in the study. Subjects were healthy non-smokers, between the ages of 18 and 22 years, who were right-handed and exhibited ipsilateral eye-dominance. All subjects provided informed consent, health history, and demographic information.

Fig. 4. Illustration of pistol shooting task

2.2 Procedures

Subjects completed two test sessions varying in stress and mental challenge. Figure 4 above illustrates the basic recording strategy of EEG monitoring during target shooting. Shooting scores were calculated as percentage accuracy by summing the point value of each target hit, dividing by the total points possible, and multiplying by 100. Concentric circles on the target range from 1 to 10, with 10 being the highest possible point total on any given shot (i.e., the bullseye).

During the low-stress condition subjects performed alone in a non-competitive state and in the absence of any time constraints to execute 40 shots. During high-stress the participants competed within a sanctioned in-house match against an opponent in an adjacent lane and under the constraint of time pressure. Specifically, the 16 participants were evenly split into two teams and the scores posted during competition were publicly displayed and entered into a composite score to determine the winning team. The order of conditions was counterbalanced.

2.3 Psychophysiological Monitoring - Signal Acquisition / Processing

High-density EEG records referenced to linked-ears were obtained via subjects wearing a 32-electrode stretch-lycra cap interfaced to high-impedance amplifiers (NeuroScan SynAmps system) supported by dedicated signal acquisition and processing software. Electrode impedance was maintained below 5Kohm, amplification was 20,000x, bandpass filtering was 1-100Hz, and a sampling rate of 512 Hz was employed. EEG records contained event markers for all shots taken. Vertical eye movements were monitored and amplified 5,000x. Electrocardiographic records (ECG) were monitored continuously throughout each session to determine parasympathetic or vagal tonus. Strength of HPA axis activity was assessed by periodic sampling of salivary cortisol during each condition. Data collection was conducted in accord with procedures published by Putnam et al. [11] to reduce risk of infection.

EEG records corrected for artifact were subjected to Fast Fourier Transform to determine regional activation (power spectral composition) and frontal asymmetry of EEG alpha power. Inter-electrode coherence was assessed between association and motor planning regions. Furthermore, the EEG records were subjected to Independent Components Analysis (ICA), as described by Contreras-Vidal and Kerick [12]**,** to: 1) conduct exploratory analyses for the reduction of movement artifact and 2) determine specified regional components (with particular emphasis on frontal and left temporal regions) and compute coherence between targeted brain regions.

3 Findings

Relative to the low-stress condition, the high-stress competitive challenge resulted in:

1. lower prefrontal asymmetry and heightened state anxiety
2. suppressed alpha power across the topography of the cortex (Figure 5)
3. increased number of cortical components clustered with a key component associated with poor performance during high stress located in the left temporal region indicative of greater cognitive complexity
4. increased alpha band coherence between left temporal and frontally located motor planning components
5. lower heart rate variability or withdrawal of vagal tone
6. heightened cortisol levels
7. degraded shooting performance

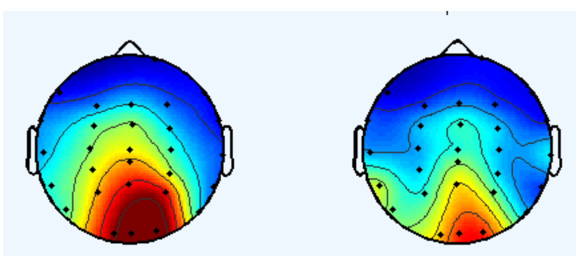


Fig. 5. Topographical EEG alpha maps during performance alone (low-stress) on left and competition (high-stress) on right at 4 seconds preceding trigger pull (Note darker shade in parietal region of low-stress condition indicating higher alpha power)

Acknowledgements. We would like to thank Michelle Costanzo, Ron Goodman, Li-Chuan Lo, Jeremy Rietschel, and Mark Saffer for technical research assistance.

References

1. Haufler, A.J., Spalding, T.W., Santa Maria, D.L., Hatfield, B.D.: Neuro-Cognitive Activity During a Self-Paced Visuospatial Task: Comparative EEG Profiles in Marksmen and Novice Shooters. Biol. Psychol. 53, 131–160 (2000)
2. Hatfield, B.D., Hillman, C.H.: The Psychophysiology of Sport: A Mechanistic Understanding of the Psychology of Superior Performance. In: Singer, R.N., Hausenblas, C.H., Janelle, C.M. (eds.) Handbook of Sport Psychology, 2nd edn., pp. 362–386. John Wiley & Sons, Inc., New York (2001)
3. Kerick, S.E., Hatfield, B.D., Allender, L.E.: Event-Related Cortical Dynamics of Soldiers During Shooting as a Function of Varied Task Demand. Aviat. Space Environ. Med. 78(5) Sect. II, B153–B164 (2007)
4. Busk, J., Galbraith, G.C.: EEG Correlates of Visual-Motor Practice in Man. Electroenceph. Clin. Neurophysiol. 38, 415–422 (1975)
5. Deeny, S., Hillman, C.H., Janelle, C.M., Hatfield, B.D.: Cortico-Cortical Communication and Superior Performance in Skilled Marksman: An EEG Coherence Analysis. J. Exerc. Sport Psychol. 25, 188–204 (2003)
6. Bear, M.F., Connors, B.W., Paradiso, M.A.: Neuroscience: Exploring the Brain. Williams & Wilkins, Baltimore (1996)
7. Davidson, R.J.: EEG measures of cerebral asymmetry: Conceptual and methodological issues. Int. J. Neurosci. 39, 71–89 (1988)
8. Davidson, R.J.: Anxiety and Affective Style: Role of Prefrontal Cortex and Amygdala. Biol. Psychiatry 51, 68–80 (2002)
9. Jackson, D.C., Mueller, C.J., Dolski, I., Dalton, K.D., Nitschke, J.B., Urry, H.L., Rosenkranz, M.A., Ryff, C.D., Singer, B.H., Davidson, R.J.: Now You Feel it, Now You Don't: Frontal Brain Electrical Asymmetry and Individual Differences in Emotion Regulation. Psychol. Sci. 14, 612–617 (2003)
10. Grafton, S.T., Hari, R., Salenius, S.: The Human Motor System. In: Toga, A.W., Mazziotta, J.C. (eds.) Brain Mapping; The Systems, pp. 331–363. Academic Press, San Diego (2000)
11. Putnam, L.E., Johnson Jr., R., Roth, W.T.: Guidelines for Reducing the Risk of Disease Transmission in the Psychophysiology Laboratory. Psychophysiol. 29, 127–141 (1992)
12. Contreras-Vidal, J.L., Kerick, S.E.: Independent Component Analysis of Dynamic Brain Responses During Visuomotor Vdaptation. Neuroimage 21, 936–945 (2004)

Long Term Repair of Learning Disability through Short-Term Reduction of CNS Inhibition

H. Craig Heller[1], Damien Colas[1], Norman F. Ruby[1], Fabian Fernandez[2], Bayarasaikhan Chuluun[1], Martina Blank[2], and Craig C. Garner[2]

[1] Departments of Biology
[2] Psychiatry and Behavioral Sciences
Stanford University, Stanford, CA 94305
hcheller@stanford.edu, cgarner@stanford.edu

Abstract. Learning disabilities are serious societal problems contributing to a loss of quality of life for affected individuals and their families. We hypothesized that the learning disability in Down Syndrome and perhaps in other neurodegenerative disorders is due to an imbalance between inhibitory and excitatory tone in the CNS. Specifically, we predicted that reduction of GABA related inhibition would improve learning. We used the TsDn65 mouse model of Down Syndrome and treated adult mice with daily doses of different GABA antagonists. Following treatments learning performance of these mice in several rodent learning tasks was indistinguishable from the performance of wild type mice, and the learning improvement lasted for months after the treatment ended. We are now exploring the mechanism of this durable neuroplastic effect and asking whether it would generalize to other learning disorders or optimize learning in wild type mice.

Keywords: GABA, picrotoxin, pentylenetetrazole, bilobilide, flumazinil, Down Syndrome, TsDn65 mice, novel object recognition.

1 Introduction

The incidence of intellectual disability is about 1-2% in western countries, and the resulting health costs, opportunity costs, and loss of productivity are in the tens of billions of dollars/year. In about three-fourths, one of several hundred single gene disorders is the cause; the remainder is due to chromosomal abnormalities, malnutrition, fetal alcohol exposure or brain injury. One of the most common, Down syndrome has been found to affect 1 of every 700 live births.

We are following up on previous work [1] that advanced and tested the hypothesis that the cognitive dysfunction in Down Syndrome is due to an excess of inhibitory tone in the CNS. It was shown that reducing inhibitory tone with non-competitive $GABA_A$ antagonists (pentylenetetrazole, bilobilide, and picrotoxin) restored learning ability in a mouse model of Down Syndrome (TsDn65 mice), and that this improvement extended months beyond the treatment regimen. The long-term efficacy of these treatments indicated that they induced a major and lasting neuroplastic effect. One

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 818–825, 2009.

indication of the neuroplastic modifications was seen is the fact that LTP in the hippocampus is extremely low in the TsDn65 mice [2, 3, 4], and was normalized by the treatment [1]. Moreover, the enhancement of LTP, like the enhancement of learning ability lasted for months after the drug treatment ended. A notable aspect of our previous work was that it was performed on adult mice. This implies that the reduction in cognitive performance caused by the triplication of over 150 syntenic human chromosome 21 genes in this Down Syndrome (DS) mouse is not permanent but can be improved to normal or near normal functionality at any age. Perhaps it can also optimize learning and memory in individuals without diagnosable disabilities.

Here we report on ongoing efforts to understand the relationship between GABA antagonism and mitigation of learning disability both short term and long term. We have more fully characterized the phenotype and the drug effects through studies of a variety of behaviors and electrophysiological measures. Our behavioral characterization has included novel object recognition, nest building, fear conditioning, and circadian organization of activity. Clear differences between DS and wild-type (WT) mice have been established, and the $GABA_A$ antagonist treatment affects all of them except poor nest building which we think is a manifestation of attention deficit. Notably, the spectral properties of REM sleep EEG, namely power in the theta band, are different in the DS and WT mice. High theta activity is also a component of alert wakefulness, and the same differences are seen during wakefulness in DS and WT mice. Treatment with the $GABA_A$ antagonists normalizes the theta activity in the DS mice.

2 Drugs, Dosing, and Behavior

The previously reported work explored the actions of non-competitive $GABA_A$ antagonists picrotoxin (1.0 mg/kg), bilobilide (5.0 mg/kg), and pentylenetetrazole (PTZ) (3.0 mg/kg) delivered orally in milk. Two behavioral tasks were used to score learning in the Ts65Dn mice: the novel object recognition task and the spontaneous alternation in a T-maze task. In both tasks, all three of these non-competitive $GABA_A$ antagonists when delivered in daily doses over 2 to 3 weeks normalized the learning behavior of the Ts65Dn mice when they were tested after the treatment regime had finished. Moreover, the improvement lasted for months after the treatment had ended. In contrast, single doses of picrotoxin and PTZ did not produce improvements in the learning behavior of the Ts65Dn mice. Thus, the treatment regimes seemed to produce gradually a semi-permanent change in the ability of the brains of the Ts65Dn mice to process and store new information. In the new work reported here, we have replicated the prior experiments using PTZ and have lowered the dose by an order of magnitude, we have extended the phenotyping of the effect of reduction of GABA inhibitory tone, and we have explored a new class of GABA antagonists.

Figure 1 shows the effects of PTZ (0.3 mg/kg ip) on performance of WT and Ts65D mice. Treatment consisted of daily injections of PTZ in saline or just saline for 16 days. The behavioral testing was done 1 week after the last injection. Novel object recognition of the Ts65Dn mice treated with PTZ was normalized.

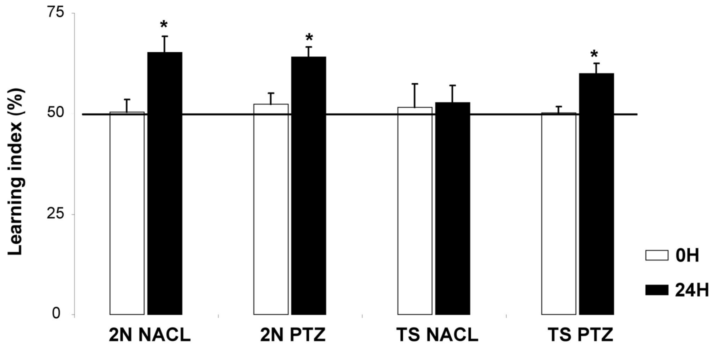


Fig. 1. Performance of WT (2N) and Ts65Dn (TS) mice 1 week after completion of 16 days of daily ip injections of PTZ at 0.3 mg/kg or saline. The open bars show that the animals do not discriminate between two objects during a 10 min learning opportunity. 24 hrs later (solid bars) all mice except the Ts65Dn animals that received vehicle can recognize a novel object from a familiar object. N = 7 to 9 animals per group.

3 Additional Behavioral Tests

We have further phenotyped the $GABA_A$ antagonism on Ts6Dn mice through fear conditioning and nest building tests. Both cued and contextual fear conditioning was improved in Ts65Dn mice treated with PTZ (3.0 mg/kg ip) for 3 weeks (Figure 2) .

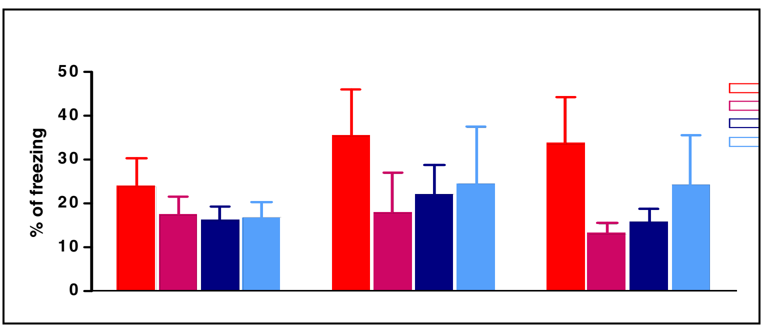


Fig. 2. Fear conditioning in PTZ treated Ts65Dn and WT mice. In each set of experiments, the left most group is the PTZ treated Ts65Dn mice, the second is Ts65Dn mice treated with vehicle, the third is WT mice treated with PTZ, and the fourth is WT mice treated with vehicle.

Mice were tested for their ability to build nests by placing a small square of compacted fiber (a Nestlet) in their cages (animals housed singly). After an hour the Nestlet (if any remained) was removed and weighed. Nest building was scored as the inverse of the proportion of the Nestlet remaining. Nests were also scored subjectively on a scale of 1 (no nest) to 5 (excellent nest). Conditions were home cage vs. new cage and thermoneutral vs. cool (to increase motivation). Ts65Dn mice showed a deficit in nest building in comparison to the WT mice in the familiar, thermoneutral environment. Both WT and Ts65Dn mice performed less well in the novel thermoneutral environment, but the deficit was much more pronounced in the Ts6Dn mice. In fact, their nest building was practically reduced to zero. Observations of their behavior via remote camera showed that their attention to the nestlets was much more fragmented. They would nudge the nestlet and then move to some other place in the cage, then return to the nestlet, and so forth. When the ambient temperature was lowered to increase motivation, both WT and Ts65Dn mice showed better nest building even in the novel environment (Figure 3). Treatment with PTZ had no effect on this aspect of the Ts65Dn phenotype, but respiridone did improve the nest building of the Ts65Dn mice (data not shown). Our tentative conclusion is that this aspect of the Ts65Dn phenotype represents an attention deficit, and this characteristic is not principally under GABAergic control.

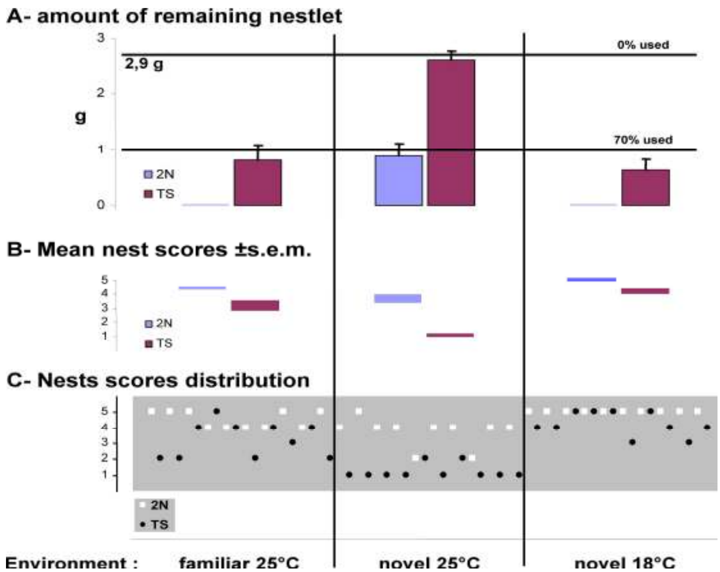


Fig. 3. Comparison of nest building behavior in Ts65Dn and WT mice

Another behavior we have studied is circadian rhythmicity. We have characterized the circadian systems of the Ts65Dn mice and compared them to WT mice, both with and without 3 weeks of PTZ treatment (3.0 mg/kg oral). The Ts65Dn mice did not show any differences in rhythm consolidation, free running characteristics in DD or LL, or ability to re-entrain to a LD cycle. The only effect of the PTZ treatment was a trend to increased activity during the active phase in the Ts65Dn mice, but not in the WT mice. Although not a significant difference, an increase in sample size would likely show this difference to be significant (Figure 4).

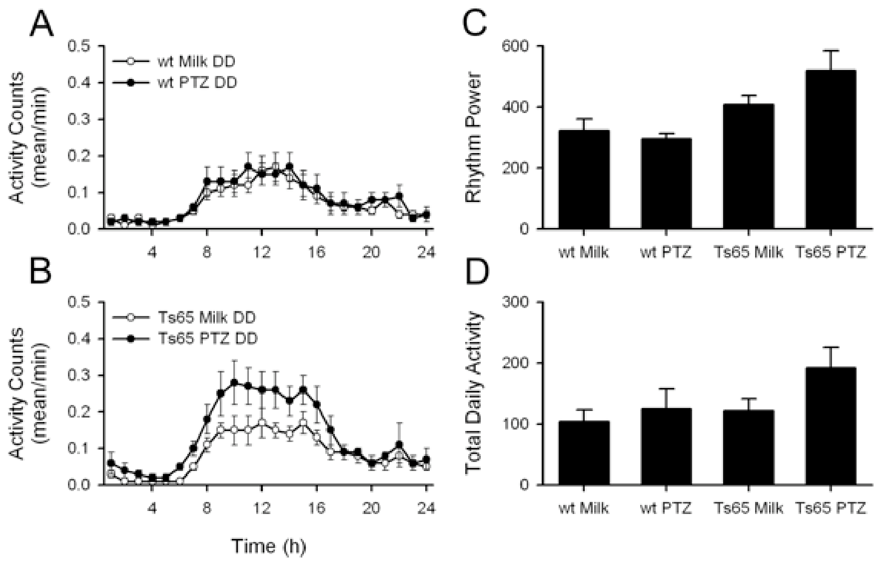


Fig. 4. Summed activity for Ts65Dn and WT mice with and without 3 weeks of treatment with daily oral doses of PTZ (3.0 mg/kg)

4 A BZD Competitive Antagonist Is also Effective in Ts65Dn Mice

Whereas all of the drugs we tested previously were non-competitive antagonists, we have recently tested a competitive BZD antagonist, Flumazenil. It had been hypothesized that release of endogenous benzodiazepines during learning helps to filter the content of consolidated memories [5]. In accordance with this idea, benzodiazepine-like immunoreactivity is reduced in the cerebral cortex, hippocampus, amygdala, and septum following brief periods of novel experience or avoidance conditioning, and injections of the competitive antagonist flumazenil in the hippocampus, amygdala, or septum improves retention in an inhibitory avoidance task [6]. Flumazenil is an attractive candidate for treatment of the learning disability of DS because it is FDA approved and it is also anti-convulsive [7]. WT and Ts65Dn mice were treated for 2 weeks with daily ip injections of either flumazenil (3 mg/kg) or vehicle. The flumazenil treated animals showed normalization of performance in the novel object recognition test performed 24 hrs after the training (Figure 5).

A remarkable result from the testing of flumazenil was that unlike PTZ or picrotoxin, a single dose of flumazenil had a positive effect on learning in the Ts65DN mice (Figure 6). The dosage in this experiment was higher (10 mg/kg), but the Ts65Dn animals showed dramatic normalization of both short-term and long-term memory.

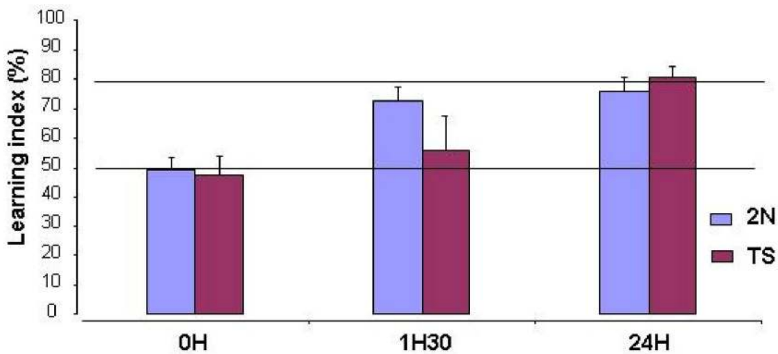


Fig. 5. Effect of 2 weeks of daily dosing of WT and Ts65Dn mice with flumazenil (3 mg/kg) or vehicle

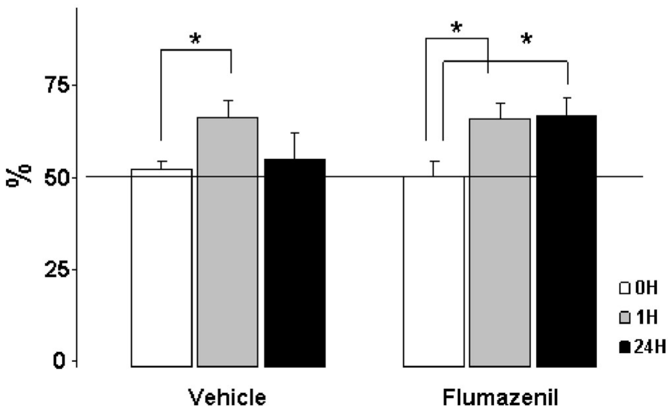


Fig. 6. Effect of a single dose of flumazenil (10 mg/kg ip) or vehicle on learning in Ts65Dn mice. N=11 in each group.

5 Electroencephalographic Studies of Effects of $GABA_A$ Antagonists

Because sleep disruptions are reported for Down syndrome patients [8], and sleep is increasingly implicated in neural plasticity, we were interested in examining the effects of $GABA_A$ antagonism on sleep characteristics in the Ts65Dn mice. Previous work [9] had only revealed subtle differences in sleep between WT and Ts65Dn mice. The Ts65Dn mice had a longer sleep latency after sleep deprivation, had less NREM sleep, and had more fragmented REM sleep during the light phase. With such subtle effects, it is not surprising that our EEG study of sleep in the mice did not show any dramatic effects of the PTZ treatment with one exception – the intensity of hippocampal theta activity in both REM sleep and waking.

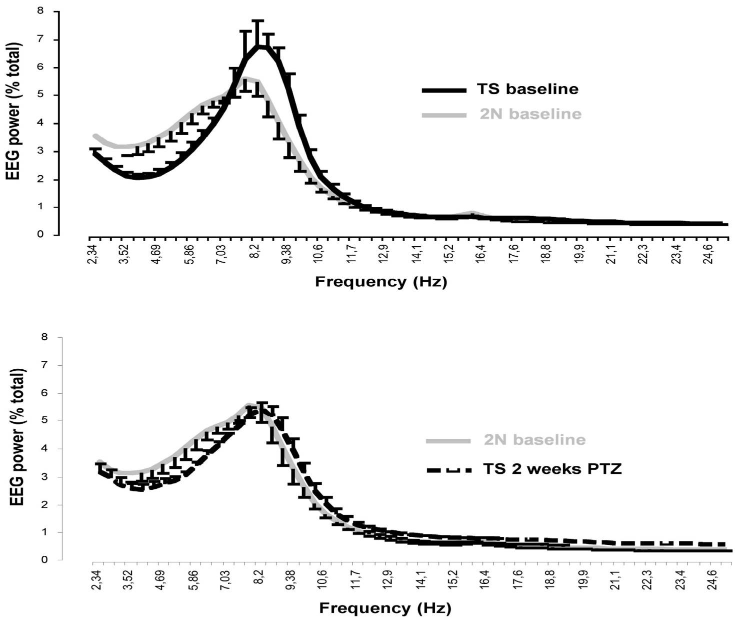


Fig. 7. Spectral analysis of the waking EEG in WT and Ts65Dn mice before and after 2 weeks of daily ip injections of PTZ (3 mg/kg) or vehicle. The sprectral power in the theta band was normalized in the Ts6Dn mice following the treatment.

Colas et al. [9] demonstrated that the Ts65Dn mice had greater spectral power in the theta band of the EEG. That observation was confirmed in the present study. Both PTZ treatment (daily ip injections 3 mg/kg for 2 weeks) (Figure 7), or acute flumazenil treatment (10 mg/kg) normalized the intensity of the EEG theta band in the Ts65Dn mice and this was true for both REM and wake.

6 Summary and Discussion

The results from our studies clearly demonstrate that negative $GABA_A$ modulators can normalize certain aspects of learning and memory in a prominent mouse model of Down syndrome. Down syndrome is the leading genetic cause of cognitive impairment in humans, and being able to mitigate that impairment would improve the quality of life for thousands of afflicted individuals and their families. Most of the drugs we have used in our studies could be candidates for human clinical trials. For decades PTZ had been in use in the U.S. for the treatment of Alzheimer's Disease, but it was taken off of the FDA list of approved drugs for lack of evidence for efficacy and not for any concerns about safety. Bilobilide is an extract of the plant *Ginko biloba* and is available as a nutriceutical. Flumazenil is an FDA approved drug for the treatment of benzodiazepine overdoses. It is likely that additional candidate drugs that would be highly selective for specific $GABA_A$ receptors exist or will be discovered. Thus, it seems to us that there would be much to gain and little risk to undertake clinical trials for pharmacological enhancement of cognition in Down syndrome. If results are positive, it is not at all unlikely that other cases of learning disabilities might also benefit from negative $GABA_A$ modulation. We do recognize that Down syndrome is a complex suite of disorders due to the large number of genes that are triplicated. $GABA_A$ modulation is not going to mitigate the vast majority of these disorders, yet cognitive ability is an extremely significant aspect of an individual's life.

One aspect of our work that is extremely provocative is the fact that short treatment with $GABA_A$ antagonists results in long term changes in neural plasticity. This effect has been shown to be expressed at the cellular level in terms of long-term potentiation [1], but the mechanisms involved remain to be elucidated.

A final question that will occur to many is whether or not normal cognition can be optimized by mild $GABA_A$ antagonism? Would this approach be more selective and effective than uses of stimulants such as caffeine and amphetamine. The results from the WT mice suggest that the answer is no, but perhaps the behavioral tasks were not difficult enough to reveal enhancement in the WT mice due to ceiling effects. Further research will have to be directed to answering this question.

Acknowledgements

This work was supported by the following: Down Syndrome Research and Treatment Foundation, The La Jeune Foundation, NIH R01-MH067752, Stanford-Coulter Translational Research Grants Program. We are grateful for the avice and help of Drs. Kevin Grimes and Daria Mochly-Rosen of the Stanford SPARK program.

References

1. Fernandez, F., Morishita, W., Zuniga, E., Nguyen, J., Blank, M., Malenka, R.C., Garner, C.C.: Pharmacotherapy for cognitive impairment in a mouse model of Down syndrome. Nat. Neurosci. 10, 411–413 (2007)
2. Siarey, H., Stoll, J., Rapoport, S.I., Galdzicki, Z.: Altered long-term potentiation in the young and old Ts65Dn mouse, a model for Down syndrome. Neuropharmacology 36, 1549–1554 (1997)
3. Costa, A.C.S., Grybko, M.J.: Deficits in hippocampal CA1 LTP induced by TBS but not HFS in the Ts65Dn mouse: A model of Down syndrome. Neuroscience Letters, 382-317-322 (2005)
4. Kleschevnikov, A.M., Belichenko, P.V., Villar, J.A.J., Epstein, C.J., Malenka, R.C., Mobley, W.C.: Hippocampal long-term potentiation suppressed by increased inhibition in the Ts65Dn mouse, a genetic model of Down syndrome. J. Neurosci. 24, 8153–8160 (2004)
5. Izquierdo, I., Da Cunha, C., Medina, J.H.: Endogenous benzodiazepine modulation of memory processes. Neurosci. Biobehav. Rev. 14, 419–424 (1990)
6. Wolfman, C., Da Cunha, C., Jerusalinsky, D., Levi de Stein, M., Viola, H., Izquierdo, I., Medina, J.H.: Habituation and inhibitory avoidance training alter brain regional levels of benzodiazepine-like molecules and are affected by intracerebral flumazenil microinjection. Brain Res. 548, 74–80 (1991)
7. Polc, P., Jahromi, S.S., Facciponte, G., Pelletier, M.R., Zhang, L., Carlen, P.L.: Benzodiazepine antagonist flumazenil reduces hippocampal epileptiform activity. Neuroreport 6, 1549–1552 (1995)
8. Hamaguchi, H., Hashimoto, T., Mori, K., Tayama, M.: Sleep in the Down syndrome. Brain Dev. 11, 399–406 (1989)
9. Colas, D., Valletta, J.S., Takimoto-Kimura, R., Nishino, S., Fujiki, N., Mobley, W.C., Mignot, E.: Sleep and EEG features in genetic models of Down syndrome. Neurobiology of Disease 30, 1–7 (2008)

Development of Sensitive, Specific, and Deployable Methods for Detecting and Discriminating mTBI and PTSD

Robin R. Johnson[1], Djordje Popvic[1,2], Deborah Perlick[3], Dennis Dyck[4], and Chris Berka[1]

[1] Advanced Brain Monitoring, Inc., Carlsbad, CA 92008
[2] University of Southern California, Los Angeles, CA 90089
[3] Mt. Sinai School of Medicine, New York, NY
[4] Washington State University, Spokane, WA
{RJohnson,Dpopvic,Chris}@b-alert.com, debbieperlick@aol.com, dyck@wsu.edu

Abstract. This paper presents a theoretical framework for the development of non-invasive methods for detection and discrimination between mild traumatic brain injury (mTBI) and post-traumatic stress disorder (PTSD). Growing use of IEDs and increased pace of multiple deployment cycles in current conflicts has lead to significant increases in exposure to risks for these conditions. Comorbidity of these conditions is common, diagnostically challenging, and controversial. Development of easy to use, deployable diagnostic tools would allow for accurate early identification and intervention. Early intervention increases the potential for positive outcomes for both the individual and their families. In addition, the appropriately designed system could be used epidemiologically to screen returning soldiers for these conditions that may otherwise not be appropriately assessed until much later, if at all. The framework presented here proposes that a wireless, portable EEG/EKG based device may be an appropriate platform upon which to develop such an assessment tool.

Keywords: Electroencephalogram (EEG), Electrocardiogram (EKG), Post-Traumatic Stress Disorder (PTSD), mild Traumatic Brain Injury (mTBI).

1 Introduction

Mild traumatic brain injury (mTBI) and post-traumatic stress disorder (PTSD) affect a growing number of our military personnel. It is thought that TBI is more common in Iraq and Afghanistan deployments than in past conflicts [1, 2]; however there is little epidemiological data available. Estimates range from 10-31% incidence rate; however these estimates are based on those with a known history of blast concussion and/or other injury that brought them to a medical unit [2, 3]. The increased use of IED (improvise explosive devices), improved body armor and improved care following injuries are often cited as the underlying reasons for the increase in TBI incidence rates [2, 4]. Despite the increase in TBI incidence rates, there is reason to think they

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 826–835, 2009.

may underestimate the actual incidence of TBI. For instance, blast injuries are the most common source of all TBI in today's military [2, 3]; since a soldier may not lose or have altered consciousness in the time frame associated with this form of injury, milder cases of TBI may remain undiagnosed. Yet the cumulative impact of incidents that individually do not meet diagnostic criteria for TBI may add up to subtle yet significant cognitive or behavioral impairments in otherwise healthy young adults actively engaged in warfare. Multiple exposures may have an additive or even synergistic effect on injury development. This gradual progression of tissue damage may go long periods of time before diagnosis, if it is detected at all.

Further complicating the diagnosis of TBI is the similarity in behavioral sequelae to those observed in PTSD. Both disorders are associated with impaired memory, inability to concentrate, inappropriate modulation of anger and other emotions [5-12]. These neuro-cognitive/neuro-affective symptoms have the greatest impact on soldiers' quality of life, family and social roles and support system maintenance, as well as ability to serve. Neurocognition and affect are behavioral outcomes of complex, multi-directional interactions between the central nervous system (CNS), endocrine system and immune system. Adding to the complexity is the ability of behavior to influence and alter each of these internal systems bi-directionally as well. With many similarities, it is possible that mTBI that develops over multiple exposures may be misdiagnosed as PTSD, leading to incomplete treatment. Therefore, it is important to develop tools that are both sensitive to these conditions, and specific in delineating across mTBI, PTSD, and cases of co-morbidity across both conditions.

The endocrine system interacts with cognition and affect primarily through the hypothalamic-pituitary-adrenal (HPA) and sympathetic-adrenal-medullary (SAM) axis. PTSD is associated with a circadian cycle dependant decrease in HPA-axis output, and the development of glucocorticoid resistance in immune cells (leading to an over-reactive inflammatory response) [13-17]. This decrease in output and reactivity may lead to memory impairments that increase with age [18]. Hypocortisolism is associated with anti-social behavior both in veterans and general populations [19-21]. There is also growing evidence that chronic stress and glucocorticoids can compromise immunity and health. In mTBI, there is some indication that hypocortisolism also develops [22] , which would result in the same down-stream consequences for social behavior and inflammation. While assessment of endocrine dysfunction may contribute to tools that are able to sensitively detect both conditions, the degree of dysfunction and/or how this dysfunction manifests across the immune and neuronal systems may aid is distinguishing across the two conditions.

Immune signaling molecules, such as the inflammatory cytokines IL-1 beta, IL-6 and TNF-alpha are also known to reduce cognitive function and impair learning and memory [23-32]. The endocrine system and immune systems interact through receptors for hormones on immune cells and for cytokines on neurons that control endocrine function [33-36]. In addition neuronal interactions also occur as a result of neuronal receptors for hormones and cytokines, neuronal production of hormones and cytokines, and direct innervation of immune and endocrine organs. Both PTSD and mTBI are associated with dysregulation of immune systems. PTSD is associated with reduced anti-inflammatory activity (IL-8)[37], over-reactive cellular immune function [38], and impaired innate immune responsivity [39]. Similarly, mTBI patients have been shown to have excessive inflammation as a secondary process associated with the injury [40, 41], however long term inflammatory processes are little studied. In contrast to PTSD, mTBI patients have reduced cellular immunity responses [42]. Both conditions have altered immune function (allowing for detection); however the disparate patterns of impairment/alteration may be useful in delineating across the conditions.

Because of the similar range of neurocognitive impairments, as well as endocrine and immune dysregulation that occur with both PTSD and mTBI, it is little wonder that controversy exists regarding the co-morbidity rates between these two conditions. Some have posited that PTSD is not possible given that TBI patients have significant memory loss associated with the injury incident [43]. However other studies have demonstrated co-morbidity up to 43% [4, 44]. Given that memory loss is less severe in injuries classified as mTBI, it is perhaps not surprising that one study found that PTSD is more common in these patients than in other TBI populations [45].

As discussed above mTBI and PTSD have many neurocognitive similarities, as well as some similarities across the endocrine, immune systems. These similarities are further complicated by the as yet unknown rate of co-morbidity across both conditions that may be from 0-43% according to various studies. The majority of research on mTBI and PTSD focuses on one, and occasionally two of these interactive systems (neuronal, cognitive, immune, endocrine), and only on 1-2 biomarkers. Advanced Brain Monitoring (ABM), has been developing a method for full neurocognitive/neurophysiological profiling (patent pending) using simultaneous EEG and EKG during administration of a basic neurocognitive testbed, in combination with blood sampling at strategic points to assess endocrine and immune activation associated with the performance in the testbed. The hypothesis is that such a multi-level approach will result in a sensitive, specific methodology for detecting and delineating across PTSD, mTBI, and co-morbid cases. This method (using ABM's B-alert wireless, portable EEG) would be both efficient and deployable into multiple environments, having been used successfully in the Mohave Desert during the summer at Twentynine Palms, in the rain at Aberdeen testing grounds, and in winter conditions in Calgary, Canada.

2 Methods

2.1 Current Assessment Methods

TBI. Currently, TBI is assessed on multiple levels in deployed environments. First, there is a strong initiative in the military to educate officers and enlisted personnel on the signs of TBI, in order to better ensure that those needing medical assessment are sent for evaluation [46]. These guidelines include looking for signs of: blurred vision, headaches, aggressive behavior, depression and cognitive issues such as trouble concentrating; and encouraging each commander to evaluate each of their soldiers regularly. More commonly, soldiers that lose consciousness at the site of engagement (i.e. at the site of an IED explosion) will be referred to medical personnel. Once referred for medical services, the soldier will be evaluated through a series of neurocognitive and imaging techniques for further referral along the chain of care. The chain of care may start with an in-country facility with basic equipment (e.g. EEG, CAT); however, TBI cannot fully be evaluated without an fMRI at this time. fMRI is not a deployable technology, however this equipment is available in country in certain cases and in the initial evacuation level facilities (i.e. such as those found in Japan and Germany).

*PTSD.*PTSD assessment and awareness program were launched simultaneous with the mTBI awareness program [46]. Signs of PTSD include: headaches, aggressive behavior, depression and cognitive issues such as trouble concentrating, and usually include flashbacks. Sleep disturbances, common in PTSD, have been suggested as playing a pathogenic role in the acute and chronic stages of the disease [47, 48]. Fragmented sleep may further contribute to the profile of neurocognitive impairments including memory loss and inability to concentrate. Assessments are typically done though extensive one-on-one evaluation sessions with a medical psychiatrist.

2.2 Proposed Assessment Methods

By measuring multiple aspects of physiology that may influence or be influenced by cognition, we may be able to build sensitive and specific mathematical models to a) detect and discriminate between those with mTBI, PTSD, and mTBI/PTSD co-morbid patients, and b) identify potential neuro-feedback intervention strategies for mTBI, PTSD, and mTBI/PTSD co-morbid patients.

Stage1- Pilot study. Initial development would require collecting EEG/EKG/ bio-marker data from a small sample size to determine feasibility. Three groups would be required: mTBI, PTSD, and healthy. A broad range of injury areas, confirmed through fMRI should be included in the mTBI group. PTSD subjects should have a confirmed diagnosis from a psychiatrist. Patients with diagnosed co-morbidity should be excluded from these data at this stage. In order to facilitate the clearest data at this stage the PTSD subjects should be screened with fMRI as well to exclude any subjects with potential undiagnosed mTBI. The minimum sample size for the pilot study is n=30/group. This will allow the feasibility of this approach to be evaluated, without investing the resources that will be required to build a truly stable mathematical model (stable models require minimum n=200, with the number of metrics in the model adding to this requirement).

The mathematical model building process would begin with exploratory descriptive discriminant analysis and/or cluster analysis that will help us narrow down a subset of physiological features predictive of each of the group, and then we shall move on to build a classifier and cross-validate it on an independent set of subjects diagnosed with mTBI, PTSD or a combination. The exact choice of the classifier and the underlying statistical model will be driven by the set of physiological features used, but we tentatively propose to use the random multinomial logit (RMNL) classifier [49]. The RMNL classifier uses forests of decision trees grown on random input vectors, the nodes split on a random subset of features, and conducts repeated multinomial logit analyses on the subsets in order to arrive at an optimal set of features while avoiding at the same time the curse of dimensionality. Given the large set and variability of parameters that will be examined, the comprehensive approach offered by the RMNL is deemed necessary.

Metrics to be collected would include: EEG, EKG, blood and saliva at regular intervals, and a computerized basic neurocognitive test battery. The neurocognitive battery will include a plethora of memory and problem solving tasks. The analytes of these metrics that may be used to build the mathematical models will include: raw EEG, PSD EEG, ABM's B-Alert and workload classifications (the EEG metrics will access brain activity as well as broad states of alertness, engagement and workload), heart rate, heart rate variability (heart rate variables will access stress and sympathetic activation), blood and/or salivary samples may be assayed for hormones, cytokines, or immune markers. Cognitive metrics will include at least reaction time, accuracy. In addition to monitoring alertness and cognitive activity, EEG characteristics of head trauma can be evaluated including automated identification of transient EEG abnormalities epileptiform ("seizure-like") patterns, such as spikes or sharp waves created when abnormal neurons synchronize and their currents summate resulting in abrupt changes from the baseline recorded as spike or sharp waves [50-52]. Continuous focal abnormalities may also be observed as alterations of ongoing EEG background activity (either attenuation or enhancement), focal slow-wave abnormalities, or periodic EEG patterns that consist of rhythmic and repetitive sharp wave or spike patterns. Each of these types of abnormalities typically is associated with underlying structural abnormality [53, 54]. EEG coherence analysis will then be applied to measure phase synchrony or shared activity between brain regions. Previous investigations have accurately identified patients with closed head injury, including assessment of severity of damage and prediction of outcomes using EEG coherence and phase analyses [52, 54, 55]. Mild to moderate TBI is often characterized by increased coherence and decreased phase in frontal and frontal-temporal regions [54], decreased power difference between anterior and posterior cortical regions [53, 56] and reduced alpha power in posterior cortical regions [56, 57].

Stage 2. If the pilot study indicates that the approach is feasible, then a full study will be conducted to meet the objective of a stable, sensitive and specific mathematical model based on non-invasive neuro-physiological metrics. The full experimental design for this study would include four groups: PTSD, mTBI, PTSD/mTBI comorbid, and healthy controls. Once again, the mTBI subjects should have a broad range of injuries, confirmed with fMRI, and the PTSD only group should have mTBI eliminated through fMRI screening. The minimal sample size from each group will be set at n=200, with adjustment based on the metrics indicated from the feasibility/pilot stage. The neurocognitive testbed will be narrowed to include only those tests that appear to discriminate across the groups in the feasibility study (i.e. a significant difference at $p < .05$). Similar mathematical model building will occur, not limited to the model developed in the feasibility study. Mathematical validation of the model will confirm stability and accuracy at this point.

Stage 3. In order to ensure that the model that has been built is sensitive, specific, and valid, a full cross validation data collection should be conducted, with sample sizes of n=200/group.

3 Results

If successful, the final product would utilize the ABM wireless, portable EEG+EKG platform (a highly cost efficient platform compared to fMRI), which is already proven deployable. Biological fluids would be collected either with filter paper based blood spot techniques or salivettes, both of which are easily used in the field for collection (although salivary samples should be kept frozen until analysis, the blood spots need only a plastic bag with a packet of desiccant). Using a suite of products that are easily deployable and easy to use in the final product will allow non-technical, non-medical staff to screen soldiers on a regular basis with a basic 20-30 minute computerized test battery. This would ideally allow soldiers that have been exposed to multiple blasts over the course of multiple deployments to be identified as developing mTBI before the injuries are exacerbated to a moderate or severe level by an additional blast. Such a solution would also allow early signs of PTSD to be identified and intervention to occur early in the process. The benefits of such a system are multi-level: 1) field deployment benefits of a cost-effective imaging based, objective assessment tool for PTSD and mTBI by non-experts; 2) improved diagnostic and treatment once in traditional care facilities (such as a Veteran's Administration hospital); 3) potential for full epidemiological screening of all returning soldiers.

While this process would require a significant investment in the initial data acquisition and model development, the final product would be highly cost-efficient, sensitive and specific; ensuring that not only soldiers were diagnosed and treated early in the injury process, but also the many persons that acquire mTBI through sports injuries and automobile accidents each year. The ABM platform is currently being successfully applied in several large multi-site clinical evaluations of patients with sleep disorders. Feasibility of conducting research comparable to the proposed studies 1-3 with samples sizes of 500-1000 has been established.

4 Discussion

mTBI and PTSD are the "signature" injuries of the Iraq and Afghanistan wars. These conditions are known to affect a large number of our soldiers. There are many indications that while both of these disorders are being diagnosed at greater levels than in previous conflicts, even more soldiers continue to go undiagnosed and untreated. There are many benefits to developing a system that would allow screening in deployed theatres as well as easy and cost effective epidemiological screening in general. While many brief screening questionnaires have been developed for both PTSD and mTBI, it is clear that a more objective system would increase sensitivity. Missed diagnoses can have significant, negative consequences for those suffering from either PTSD or mTBI.

While many soldiers with mTBI are able to return to active duty either with or without rehabilitation therapy [58], undiagnosed mTBI can result in career ending consequences. As noted earlier, mTBI is associated with conduct-disorder-like behavior, and cognitive function impairments. Conduct-disorder-like behavior may lead to inappropriate aggression, physical or verbal confrontations with peers or superiors, domestic abuse, and potentially suicidal behavior. Cognitive function impairment may lead to the soldier being incapable of meeting the expectations of their command, making inappropriate decisions either during training or deployments that could lead to injury for themselves or others. In addition to the consequences while in service, these issues may impact job performance, the ability to acquire employment, and other long term consequences. Perhaps the most important undiagnosed PTSD can have similar consequences to military-related performance, personal life, and post-military development.

An additional application for the device proposed may include application as a neurofeedback rehabilitation device to provide functional therapeutics for mTBI and PTSD patients as well. Because the incidence of PTSD and mTBI are growing each year as conflicts in Iraq and Afghanistan continue, and the consequences the development of devices, methodologies, and systems that identify all of those that should be further diagnosed and/or treated has become essential.

References

1. Jontz, S.: TBI during wartime: the Afghanistan and Iraq Experience; More head injuries in Afghanistan, Iraq push improvements in protective gear (2006), `http://www.stripes.com/article.asp?section=104&article=20155` (cited February 3, 2009)
2. Warden, D.: Military TBI during the Iraq and Afghanistan wars. J. Head Trauma Rehabil. 21(5), 398–402 (2006)
3. Hoge, C.W., et al.: Mild traumatic brain injury in U.S. Soldiers returning from Iraq. N. Engl. J. Med. 358(5), 453–463 (2008)
4. Schneiderman, A.I., Braver, E.R., Kang, H.K.: Understanding sequelae of injury mechanisms and mild traumatic brain injury incurred during the conflicts in Iraq and Afghanistan: persistent postconcussive symptoms and posttraumatic stress disorder. Am. J. Epidemiol. 167(12), 1446–1452 (2008)
5. Frewen, P.A., Lanius, R.A.: Toward a psychobiology of posttraumatic self-dysregulation: reexperiencing, hyperarousal, dissociation, and emotional numbing. Ann. N. Y. Acad. Sci. 1071, 110–124 (2006)
6. Hoge, C.W., et al.: Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. N. Engl. J. Med. 351(1), 13–22 (2004)
7. Killgore, W.D., et al.: Post-combat invincibility: violent combat experiences are associated with increased risk-taking propensity following deployment. J. Psychiatr Res. 42(13), 1112–1121 (2008)
8. Johnsen, G.E., Asbjornsen, A.E.: Consistent impaired verbal memory in PTSD: A meta-analysis. J. Affect Disord. 111(1), 74–82 (2008)
9. Horner, M.D., Hamner, M.B.: Neurocognitive functioning in posttraumatic stress disorder. Neuropsychol Rev. 12(1), 15–30 (2002)
10. Breslau, N.: Outcomes of posttraumatic stress disorder. J. Clin. Psychiatry 62(Suppl. 17), 55–59 (2001)
11. Niemann, H., Ruff, R.M., Baser, C.A.: Computer-assisted attention retraining in head-injured individuals: a controlled efficacy study of an outpatient program. J. Consult Clin. Psychol. 58(6), 811–817 (1990)
12. Niemann, H., Ruff, R.M., Kramer, J.H.: An attempt towards differentiating attentional deficits in traumatic brain injury. Neuropsychol Rev. 6(1), 11–46 (1996)
13. Rohleder, N., et al.: Hypocortisolism and increased glucocorticoid sensitivity of pro-Inflammatory cytokine production in Bosnian war refugees with posttraumatic stress disorder. Biol. Psychiatry 55(7), 745–751 (2004)
14. de Kloet, C.S., et al.: Leukocyte glucocorticoid receptor expression and immunoregulation in veterans with and without post-traumatic stress disorder. Mol. Psychiatry 12(5), 443–453 (2007)

15. Baker, D.G., et al.: Plasma and cerebrospinal fluid interleukin-6 concentrations in posttraumatic stress disorder. Neuroimmunomodulation 9(4), 209–217 (2001)
16. Vidovic, A., et al.: Circulating lymphocyte subsets, natural killer cell cytotoxicity, and components of hypothalamic-pituitary-adrenal axis in Croatian war veterans with posttraumatic stress disorder: cross-sectional study. Croat Med. J. 48(2), 198–206 (2007)
17. Belda, X., et al.: Exposure to severe stressors causes long-lasting dysregulation of resting and stress-induced activation of the hypothalamic-pituitary-adrenal axis. Ann. N. Y. Acad. Sci. 1148, 165–173 (2008)
18. Yehuda, R., et al.: Enhanced effects of cortisol administration on episodic and working memory in aging veterans with PTSD. Neuropsychopharmacology 32(12), 2581–2591 (2007)
19. Mason, J.W., et al.: Marked lability in urinary cortisol levels in subgroups of combat veterans with posttraumatic stress disorder during an intensive exposure treatment program. Psychosom Med. 64(2), 238–246 (2002)
20. Pajer, K., Rabin, B., Gardner, W.: Increased IgG 3:4 ratios in adolescent antisocial females: evidence of Th1/Th2 imbalance? Brain Behav. Immun. 16(6), 747–756 (2002)
21. Thomsen, A.F., et al.: The risk of affective disorders in patients with adrenocortical insufficiency. Psychoneuroendocrinology 31(5), 614–622 (2006)
22. Wachter, D., et al.: Pituitary insufficiency after traumatic brain injury. J. Clin. Neurosci. 16(2), 202–208 (2009)
23. Shapira-Lichter, I., et al.: Cytokines and cholinergic signals co-modulate surgical stress-induced changes in mood and memory. Brain Behav Immun. 22(3), 388–398 (2008)
24. Maier, S.F., Watkins, L.R.: Immune-to-central nervous system communication and its role in modulating pain and cognition: Implications for cancer and cancer treatment. Brain Behav Immun. 17(Suppl. 1), S125–S131 (2003)
25. Baune, B.T., et al.: Association between genetic variants of IL-1beta, IL-6 and TNF-alpha cytokines and cognitive performance in the elderly general population of the MEMO-study. Psychoneuroendocrinology 33(1), 68–76 (2008)
26. Chen, J., et al.: Neuroinflammation and disruption in working memory in aged mice after acute stimulation of the peripheral innate immune system. Brain Behav Immun. 22(3), 301–311 (2008)
27. Meyers, C.A., Albitar, M., Estey, E.: Cognitive impairment, fatigue, and cytokine levels in patients with acute myelogenous leukemia or myelodysplastic syndrome. Cancer, 2005 104(4), 788–793 (2005)
28. Banks, W.A., Farr, S.A., Morley, J.E.: Entry of blood-borne cytokines into the central nervous system: effects on cognitive processes. Neuroimmunomodulation 10(6), 319–327 (2002)
29. Reichenberg, A., et al.: Cytokine-associated emotional and cognitive disturbances in humans. Arch Gen. Psychiatry 58(5), 445–452 (2001)
30. Rosczyk, H.A., Sparkman, N.L., Johnson, R.W.: Neuroinflammation and cognitive function in aged mice following minor surgery. Exp. Gerontol. (2008)
31. Larson, S.J., Hartle, K.D., Ivanco, T.L.: Acute administration of interleukin-1beta disrupts motor learning. Behav Neurosci. 121(6), 1415–1420 (2007)
32. Sparkman, N.L., et al.: Interleukin-6 facilitates lipopolysaccharide-induced disruption in working memory and expression of other proinflammatory cytokines in hippocampal neuronal cell layers. J. Neurosci. 26(42), 10709–10716 (2006)
33. Marshall, L., Born, J.: Brain-immune interactions in sleep. Int. Rev. Neurobiol. 52, 93–131 (2002)

34. Eskandari, F., Sternberg, E.M.: Neural-immune interactions in health and disease. Ann. N. Y. Acad. Sci. 966, 20–27 (2002)
35. Ballieux, R.E.: The mind and the immune system. Theor. Med. 15(4), 387–395 (1994)
36. Dantzer, R., Kelley, K.W.: Twenty years of research on cytokine-induced sickness behavior. Brain Behav Immun. 21(2), 153–160 (2007)
37. Song, Y., et al.: Disturbance of serum interleukin-2 and interleukin-8 levels in posttraumatic and non-posttraumatic stress disorder earthquake survivors in northern China. Neuroimmunomodulation 14(5), 248–254 (2007)
38. Altemus, M., Dhabhar, F.S., Yang, R.: Immune function in PTSD. Ann. N. Y. Acad. Sci. 1071, 167–183 (2006)
39. Ironson, G., et al.: Posttraumatic stress symptoms, intrusive thoughts, loss, and immune function after Hurricane Andrew. Psychosom Med. 59(2), 128–141 (1997)
40. Morganti-Kossmann, M.C., et al.: Modulation of immune response by head injury. Injury 38(12), 1392–1400 (2007)
41. Lenzlinger, P.M., et al.: The duality of the inflammatory response to traumatic brain injury. Mol. Neurobiol. 24(1-3), 169–181 (2001)
42. Mazzeo, A.T., et al.: Severe human traumatic brain injury, but not cyclosporin a treatment, depresses activated T lymphocytes early after injury. J. Neurotrauma 23(6), 962–975 (2006)
43. Sbordone, R.J., Liter, J.C.: Mild traumatic brain injury does not produce post-traumatic stress disorder. Brain Inj. 9(4), 405–412 (1995)
44. Kennedy, J.E., et al.: Posttraumatic stress disorder and posttraumatic stress disorder-like symptoms and mild traumatic brain injury. J. Rehabil. Res. Dev. 44(7), 895–920 (2007)
45. Klein, E., Caspi, Y., Gil, S.: The relation between memory of the traumatic event and PTSD: evidence from studies of traumatic brain injury. Can. J. Psychiatry 48(1), 28–33 (2003)
46. Reed, C.: PTSD and TBI awareness programs launched (2007), `http://www.stripes.com/article.asp?section=104&article=50046` (cited February 5, 2009)
47. Hartmann, E.: A note on the nightmare. Int. Psychiatry Clin. 7(2), 192–197 (1970)
48. Ross, R.J., et al.: Sleep disturbance as the hallmark of posttraumatic stress disorder. Am. J. Psychiatry 146(6), 697–707 (1989)
49. Prinzie, A., Van den Poel, D.: Random forests for multiclass classification: Random multinomial logit. Expert Systems with Applications 34(3), 1721–1732 (2008)
50. Hoffman, D.A., et al.: Diagnosis and treatment of closed head injury. Journal of Neurotherapy 1(1), 14–21 (1995)
51. Johnstone, J., Thatcher, R.W.: Quantitative EEG analysis and rehabilitation issues in mild traumatic brain injury. J. Insur. Med. 23(4), 228–232 (1991)
52. Pointinger, H., et al.: Electroencephalography in primary diagnosis of mild head trauma. Brain Inj. 16(9), 799–805 (2002)
53. Thatcher, R.W., et al.: Biophysical linkage between MRI and EEG amplitude in closed head injury. Neuroimage 7(4 Pt. 1), 352–367 (1998)
54. Thatcher, R.W., et al.: Biophysical linkage between MRI and EEG coherence in closed head injury. Neuroimage 8(4), 307–326 (1998)
55. Thatcher, R.W., Krause, P.J., Hrybyk, M.: Cortico-cortical associations and EEG coherence: a two-compartmental model. Electroencephalogr Clin. Neurophysiol. 64(2), 123–143 (1986)

56. Thatcher, R.W., et al.: An EEG severity index of traumatic brain injury. J. Neuropsychiatry Clin. Neurosci. 13(1), 77–87 (2001)
57. Thatcher, R.W., et al.: EEG discriminant analyses of mild head trauma. Electroencephalogr Clin. Neurophysiol. 73(2), 94–106 (1989)
58. Gottshall, K.R., et al.: To investigate the influence of acute vestibular impairment following mild traumatic brain injury on subsequent ability to remain on activity duty 12 months later. Mil. Med. 172(8), 852–857 (2007)

Physiologically Driven Rehabilitation Using Virtual Reality

Angela M. Salva[1], Antonio J. Alban[1], Mark D. Wiederhold[2], Brenda K. Wiederhold[2], and Lingjun Kong[2]

[1] The Virtual Reality Medical Center, Orlando, Florida, USA
[2] The Virtual Reality Medical Center, San Diego, California, USA

Abstract. Creating a platform that allows the fusion of real-time physiological measurements and virtual reality (VR) simulation will greatly improve present human-computer interaction, adaptive displays, military training, and anxiety therapy. The Virtual Reality Medical Center (VRMC) has developed a physiologically-driven rehabilitation platform that correctly assesses user anxiety levels based on multiple real time physiological measures, determines the optimal level of physiological arousal for each individual user, and automates the virtual simulation to the proper intensity for each user. Additionally, VRMC collaborates with UCF to develop novel, state-of-the-art sensors to be integrated within the platform that are capable of measuring electrocardiogram, (EEG), skin conductance, gait, and pupillometry. In Phase I VRMC developed a capability to monitor, fuse, and evaluate physiological measures (heart rate, skin conductance, skin temperature, and respiration) in real time to assess user anxiety levels. The physiological data collected will be used to assess user anxiety levels in real time as neutral, low, or high with 90% accuracy and to determine the optimal level of physiological arousal for each individual user.

Keywords: physiological measurement, stroke, traumatic brain injury, cerebrovascular accident, rehabilitation, cognitive rehabilitation, simulation, mixed reality.

1 Introduction

Modern human-computer interaction (HCI) development has recently been focusing on creating user-centered applications that adapt to the mental, or cognitive, state of the user. These systems commonly measure EEG [2, 15, Wilson & Russel, 2004), pupillometry [16], and cardiac function (Liddle et al., 2005) to evaluate user mental load in real time, and adapt the displays accordingly. If cognitive functioning is low, the display adapts to engage the user; if the user is overwhelmed, the display lessens its demands or stimuli to allow the user to focus. Some systems train users to control their physiology, e.g., brainwaves as measured on EEG, to control displays, such as driving simulators for those who have acquired brain injury (Lew et al., 2005) or communication devices for people who are completely paralyzed [1]. These human-in-the-loop systems are extremely valuable in training and rehabilitation of these populations.

D.D. Schmorrow et al. (Eds.): Augmented Cognition, HCII 2009, LNAI 5638, pp. 836–845, 2009.

There are many challenges, though, in creating such a system. First, the system needs a battery of strategies to recognize user mental status; that is, developing techniques (e.g., data fusion) to determine what the physiological input from the user means – is the user bored, overwhelmed, or distracted? Once the system can recognize the mental state of the user, developers must train the system to decide how the input will determine the output, and how the output will relate to the user. Next, developers must ensure the accuracy of the input recognition and output decisions the system is making, as data fusion and processing in real-time places increased demands on the software. Finally, once the system accurately assesses the input from the user, it must learn to correctly adapt its display based on the input to create a successful human-in-the-loop system.

Mixed Reality (MR) is a simulation technology that blends virtual reality with physical reality into a seamless landscape. The advantage of MR is that it creates an altered or augmented reality without losing the benefits of the physical setting - touch, smell, hearing, taste, and visual contact with other humans. The MRRS will enable Cerebrovascular (CVA) patients to receive physical and cognitive rehabilitation both in the therapist's office and at home. CVA patients include those impaired from a stroke or traumatic brain injury (TBI). The Virtual Reality Medical Center (VRMC) has developed the MRRS to provide an interactive, engaging rehabilitation tool for these patients.

Approximately 700,000 Americans are affected by stroke annually, costing an estimated $62 billion in 2008. A recent RAND survey found that 19.5% (over 320,000) of service members may have experienced at least a mild TBI while deployed. Multiple re-deployments, unprecedented in this all-volunteer U.S. military, may compound the risk for physical and psychological injuries, potentially resulting in more severe and chronic mental health problems. In 2008, the annual level of suicides among soldiers was the highest it has been since the Pentagon began tracking the rate 28 years ago. Expanded development of the MRRS adds the ability to perform cognitive assessment for CVA and stress injuries, including Post Traumatic Stress Disorder (PTSD), by capturing and analyzing the patient's reactions and performance while in a controlled environment. The ability to detect mental health issues with the MRRS before and after deployments could save lives.

2 Review of Literature

Much research has been done examining the role physiology has in "peak performance" from athletic to military training. Researchers have determined that most individuals are unaware of the effect their thoughts have on their physiology, and in turn, the effects their physiology has on their performance or execution of a task [3, 5, 10, 17].

The Yerkes-Dodson curve, illustrated in Figure 1, shows the inverted U-curve of anxiety's relationship with performance. Physiologically, there is an optimal level of anxiety/arousal that influences individual's performance efficiency. As anxiety increases, performance efficiency improves and reaches an optimal point. As shown by the curve, if a person becomes over-anxious or complacent, his or her performance efficiency will suffer. Low anxiety does not allow someone to become invested in

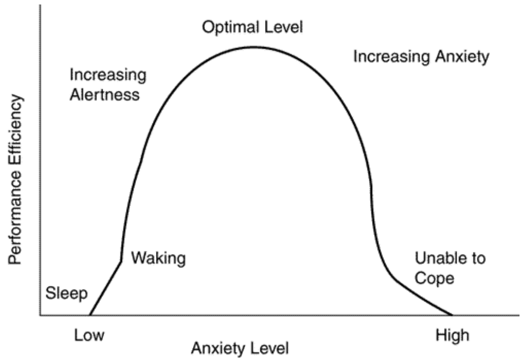


Fig. 1. Yerkes-Dodson curve showing the relationship between anxiety (arousal) and performance

executing an activity, while high anxiety leaves a person unable to cope in a stressful situation. Neither extreme is conducive to optimal efficiency in executing a task.

A CVA patient experiencing phobic stimuli may be at the high-anxiety end of the curve. These individuals are left unable to cope or execute tasks when faced with stimuli, and may even experience panic attacks or physical discomfort. These types of reactions are completely debilitating when trying to perform activities of daily living. It is necessary for patients experiencing this discomfort to understand the role physiology has in their reaction to stimuli and how to manage their physiology when faced with a provoking situation. Typical physiological reactions of people experiencing anxiety include:

- *Increased heart rate (HR).* Heart rate has been considered a particularly strong measure of anxiety [8, Nesse et al., 1985].
- *Decreased skin resistance (SR).* In 1907, Carl Jung discovered that skin resistance (SR, which decreases as sweat gland activity increases) was a means to objectify emotional tones previously thought to be invisible. Skin resistance, unlike electromyography (EMG) and skin temperature, tends to reflect mental events more quickly and with more resolution than other physiological measures [9]. Baseline levels of SR vary widely by individual so percentage change from baseline is normally measured rather than absolute value [13].

– *Drop in skin temperature.* Circulation slows in the extremities during stress, causing skin temperature to drop. Although change in skin temperature is less sensitive than and temporally lags changes in heart rate, its response curves are similar [6, 11].

– *Poor respiration.* Phobics typically show increased breaths-per-minute and less respiratory sinus arrhythmia (RSA) than non-phobics during exposure to phobic stimuli [14], and patients with anxiety disorders exhibit decreased RSA in general [4, 8].

While these physiological changes are often measured during anxiety or phobia therapy, whoever is monitoring the physiology must interpret the significance of each measure in relation to the patient, as well as in relation to the separate physiological signals. This can lead to discrepancies in treatment. While physiological monitoring and biofeedback greatly improve therapy outcomes, these techniques could be greatly improved upon by technological advancements, specifically data fusion and signal processing to interpret anxiety levels in patients. Having a network that evaluates patients' anxiety levels allows therapists to focus on individualizing treatment by teaching patients how their physiology is affecting their everyday living, as well as techniques to cope and manage physiology to improve their cognition and lessen their anxiety.

3 Method/Approach

In an effort to improve cognitive deficits and diminish abnormal behaviors caused by brain trauma, VRMC, partnered with the Media Convergence Lab (MCL) at the University of Central Florida (UCF) Institute for Simulation and Training (IST), to create a haptics-enhanced true 3D stereo mixed reality system especially designed to stimulate and improve cognitive functions in warfighters that suffer from Traumatic Brain Injury (TBI) and CVA patients.

TBI is the most common combat-related injury. It often results in disturbances of attention, memory, and executive function; moderate to severe cases can cause seizures. Sixty percent of troops who survive external injuries from bomb blasts, the leading cause of death in Operation Iraqi Freedom, could also have brain injury. While there are potential drug-based candidates for neuroprotection of brain injuries, comprehensive-holistic neuropsychological rehabilitation that attempts to address multiple cognitive deficits seems to be effective for the remediation of attention deficits and memory impairments after TBI; however, such intensive daily treatment within hospitals would no doubt be costly and has only been shown to be effective in mild memory impairments. Finding alternative, cost-effective ways to rehabilitate soldiers would help save the military and government a significant amount of resources.

Recent advances in communication and visualization technologies are resulting in the ability for a mobile user to effectively "browse" a physical environment and obtain site-specific information or access representations of real-time data about their immediate location. Our research has focused on combining mobile multimedia, virtual reality, and wearable computing technologies, to create systems that provide MR experiences. MR is a type of virtual reality that combines real and computer generated images to create an augmented reality. The existing system also provides multisensory feedback, including auditory and tactile feedback.

The MRRS incorporates scenarios, under the direction of the therapists, that stress the importance of activities of daily life (ADLs) and seek to improve patients' independence by retraining them in routine activities necessary for daily living. The human factors study examined the ergonomics of the system setup and the validity of mixed reality. The MRRS was evaluated by test participants. All participants were able to complete the 9 minutes 41 seconds scenario. Every participant was outfitted with biofeedback equipment to measure physiological effects of the experience. Fourteen healthy participants (six males and eight females) were enrolled in this study. Participants were recruited at the University of Central Florida. Participant ages ranged from 18 years of age to 63 years of age. They varied in their experience and familiarity with video games and mixed reality.

Participant physiological measurements were monitored by the J&J Engineering's I-330-C2-system. This system measured the participant's heart rate, skin conductance, skin temperature, respiratory effort, and breaths per minute. Participants also filled out self-report questionnaires that included: Parent's Modified Simulator Sickness Questionnaire consisting of 38 questions, the seven item Presence Questionnaire, the State-Trait Anxiety Inventory, the Tellegen Absorption Scale, and the Dissociative Experiences Scale.

Additionally, participants provided subjective feedback in a structured interview. This included verbal rankings on a Likert scale of 0 to 10 for level of enjoyment, level of comfort, and ease of use, where 0 was none at all and 10 was most enjoyable, comfortable, and/or easy to use. Participants also ranked their accuracy in completing the tasks assigned and overall performance on a 0-100% scale, in which 100% resembled the highest accuracy and/or most proficient performance. Each participant would also provide subjective feedback regarding the functionality and physicality of the system. With negative feedback, the participant provided possible ways to improve the system.

Participants were first required to provide a signed informed consent form following a discussion of the possible risks with the consent administrator. Participants were asked to complete a set of two pre-questionnaires. Next the participants were familiarized with the MRRS setup. At this time, their role in the scenario and the three tasks they would be responsible for were explained. Spatial audio tests and scanner instruction tests were also performed. Following these instructions, all participants were then fitted with the head-mounted display (HMD). Participants then completed a few of the assigned in-game tasks to familiarize themselves with the equipment. Then the participants were asked to relax, close their eyes and concentrate on their breathing for five minutes while a physiological baseline was established. Finally the participants completed the three tasks assigned to them in the supply depot scenario.

Following the scenario, participants were interviewed by a research assistant. Participants answered questions that dealt with comfort, ease of use, performance, accuracy, and replay value. Three self-evaluated questionnaires were collected at the end of the session.

4 Results

Physiological measures were collected from all 14 participants that participated in this study. Table 1 below depicts the average physiological measure of heart rate, skin conductance, respiratory effort, breaths per minute, and temperature from the baseline and the scenario. The results show increases over baseline readings in heart rate, skin conductance, and breaths per minute after participants had executed the scenario.

All participants filled out a Simulator Sickness Questionnaire following the usability testing. This questionnaire presents users with several symptoms, such as Headache and Nausea, which may result from interacting in a virtual environment. Participants ranked the symptoms on a scale of 0 to 3, indicating if symptoms were

Table 1. Average Physiological Measures

Situation	Physiological Measures				
	HR	SC	BPM	RESP	TEMP
Baseline	82.03086	7.843363	10.44903	541.589	85.55564
Scenario	92.36581	15.71259	15.36546	407.133	81.25739

(0) absent, (1) slight, (2) moderate, or (3) severe. The overall average score for simulator sickness during exploration of the virtual environment was 0.283. Difficulty focusing and blurred vision were the two most experienced symptoms. Two out of the 14 participants did not feel any symptoms at all.

According to user feedback, the MRRS achieved an average enjoyment rating of 7.14. Participants gave the level of comfort an average 7.00 rating, while the ease of use generated an 8.82 rating. Based on correlations calculated using a Pearson correlation coefficient, the variables level of comfort and level of enjoyment is positively correlated. Figure 2 depicts the level of enjoyment versus the level of comfort.

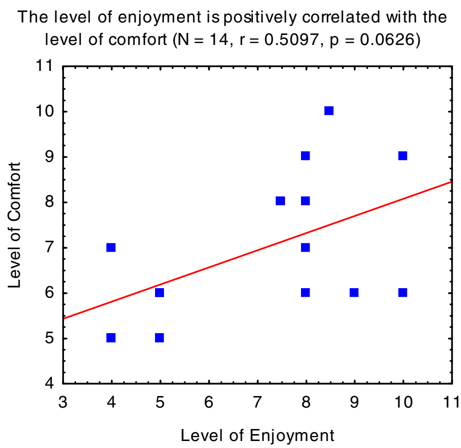


Fig. 2. Level of Enjoyment Versus the Level of Comfort

In addition to the Simulator Sickness Questionnaire, participants were asked to complete the Presence Questionnaire. Presence was scored on a scale of 1-7. The overall average score for the Presence Questionnaire was 5.35. When correlated with the ease of use, presence is found to be positively correlated with the system's ease of use. Figure 3 depicts the ease of use versus the presence score.

Participants considered their accuracy and overall performance to be 74.07% and 70.77%, respectively. The perceived overall performance rating was positively correlated to the level of comfort. Figure 4 depicts the level of comfort versus overall performance rating.

Subjective user feedback revealed that 11 out of the 14 participants encountered difficulties reading any type of text through the head-mount device (HMD). Five out of the 14 participants mentioned the latency in the HMD. Participants felt that there

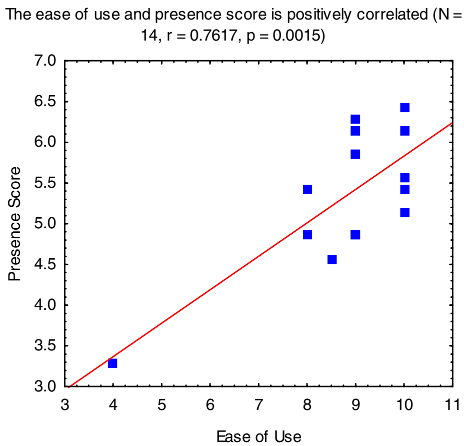


Fig. 3. Ease of Use Versus the Presence Score

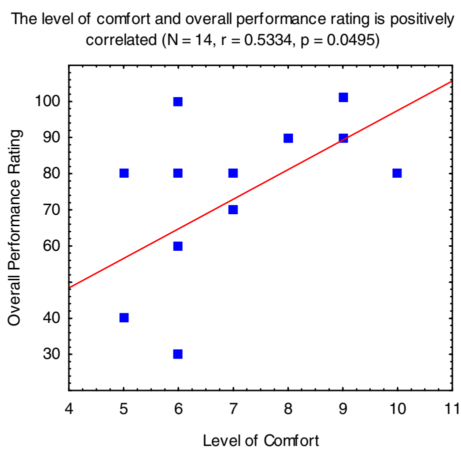


Fig. 4. Level of Comfort versus Overall Performance Rating

was a lag in the display when moving the head. Ten out of the 14 participants felt the audio was realistic and presented no flaws. Among the 4 participants that claimed the audio was flawed or stated that it was too loud, which covered the sound of the trucks parking in the loading docks. Thus, this prevented them from accomplishing one of their tasks.

The results provide evidence that this MRRS presents a user friendly interface, immersive virtual interaction, and an enjoyable experience. The participants gave themselves high ratings in accuracy and overall performance, indicating that they were engaged and participating fully in the scenario. The correlation between the ease of use and presence score indicates that the user's level of immersion can ultimately be improved by making the virtual environment more user friendly. This would include making the controls easier to function and the virtual environment easier to navigate. Also, tasks must avoid complication by remaining simple without several steps.

The presence scores suggest that the MRRS commands an immersive experience. Based on subjective user feedback, the scenario provided the participants with enough activities to keep them busy. Many felt the amount of tasks were sufficient despite the constant audio and visual distractions. The scenario displayed higher physiological readings in categories of heart rate, skin conductance and breaths per minute. Generally, an increase in heart rate, skin conductance, and breaths per minute indicates more activity and arousal. Temperature can be used as a physiological indicator of distress and anxiety, as temperature decreases through stress induced vasoconstriction and increases through vasodilation caused by relaxation. Thus, a higher average temperature in the baseline suggests a reduced level of discomfort and anxiety.

5 Discussion

The use of novel, miniaturized, portable sensors and electrodes improves patient acceptance and also broadens the field of possible applications of physiologically-driven MRRS. In addition, the new sensors may lead to completely unobtrusive methods for physiological data collection, depending on the way MRRS is implemented. The collection and measurement of multiple emotional conditions will help standardize and enhance rehabilitation in general. Improved sensor recognition of emotions and behavior from speech, EEG, facial expression, and other natural reactions will also provide the patient with more accurate audio, visual and haptic feedback.

The MRRS has many capabilities and great potential. The current state of the system could easily be augmented to accommodate various levels of cognitive functioning. For example, a scenario involving making cereal could have various levels of complexity. The easiest level may have all the ingredients open or ready for the user to prepare cereal. An intermediate level may require some prior preparation, such as opening the milk carton or cereal box. To further challenge the patient, he or she may have to find and retrieve one or more of the necessary components from around the virtual kitchen: a bowl from cabinets, a spoon from the drawer, milk from the refrigerator, and cereal box from the pantry. Other scenarios of task training will have similar levels of difficulty.

6 Conclusions and Recommendations

There are many reasons why MR applications may be effective for rehabilitation. First, MR, like VR, is an interactive, experiential medium. In the same way that children and teenagers intuitively grasp computers, MR users become directly engaged with the effects of the mixed reality experience. In addition, MR creates a safe setting where patients can explore and act without feeling threatened [7]. Patients can make mistakes without fear of dangerous, real, or humiliating consequences. Moreover, unlike human trainers, computers are infinitely patient and consistent. In cognitive rehabilitation, MR can be manipulated in ways that the real world cannot. For example, MR can convey rules and abstract concepts without the use of language or symbols for patients with little or no grasp of language.

MR creates a safe, controlled environment for repetitive practice, which is crucial in learning tasks, while providing immediate, real-time feedback about performance. Because of its interactive nature, MR can increase compliance by making the experience fun. While these technologies will have immediate benefit for CVA patients, their development will also serve to catalyze improvement and change within clinical rehabilitation at large. MR may indeed help create a more enjoyable and effective method of rehabilitating patients with brain injuries than the current paradigm.

References

1. Birbaumer, N., Kübler, A., Ghanayim, N., Hinterberger, T., Perelmouter, J., Kaiser, J., Iversen, I., Kotchoubey, B., Neumann, N., Flor, H.: The Thought Translation Device (TTD) for Completely Paralyzed Patients. IEEE Transactions on Rehabilitation Engineering 8(2), 190–193 (2000)
2. Chen, D., Vertegaal, R.: Using mental load for managing interruptions in physiologically attentive user interfaces. In: Late breaking results paper, Conference on Human Factors in Computing Systems (CHI 2004), Vienna, Austria, April 24-29, pp. 1513–1516 (2004)
3. Esteve-Lanao, J., San Juan, A.F., Earnest, C.P., Foster, C., Lucia, A.: How do endurance runners actually train? Relationship with competition performance. Medicine and Science in Sports and Exercise 37(3), 496–504 (2005)
4. Kollai, M., Kollai, B.: Cardiac vagal tone in generalised anxiety disorder. British Journal of Psychiatry 161, 831–835 (1992)
5. Lieberman, H.R., Bathalon, G.P., Falco, C.M., Kramer, F.M., Morgan III, C.A., Niro, P.: Severe decrements in cognition function and mood induced by sleep loss, heat, dehydration, and undernutrition during simulated combat. Biological Psychiatry 57(4), 422–429 (2005)
6. Meehan, M., Insko, B., Whitton, M., Brooks Jr., F.P.: Physiological measures of presence in stressful virtual environments. In: Association for Computing Machinery Transactions on Graphics, Special Issue: Proceedings of ACM SIGGRAPH 2002, vol. 21(3), pp. 645–652 (2002)
7. Riva, G.: Virtual Reality in Psychotherapy: Review. CyberPsychology & Behavior 8(3), 220–230 (2005)
8. Roth, W.T., Telch, M.J., Taylor, C.B., Sachitano, J.A., Gallen, C.C., Kopell, M.L., McClenahan, K.L., Agras, W.S., Pfefferbaum, A.: Autonomic characteristics of agoraphobia with panic attacks. Biological Psychiatry 21(12), 1133–1154 (1986)
9. Schwartz, M.S.: Biofeedback: A Practitioner's Guide. Guilford Press, New York (1995)
10. Seiler, K.S., Kjerland, G.O.: Quantifying training intensity distribution in elite endurance athletes: Is there evidence for an 'optimal' distribution? Scandinavian Journal of Medicine & Science in Sports (in press)
11. Slonim, N.B.: Environmental Physiology. The C.V. Mosby Company, Saint Louis (1974)
12. Walshe, D.G., Lewis, E.J., Kim, S.I., O'Sullivan, K., Wiederhold, B.K.: Exploring the use of computer games and virtual reality in exposure therapy for fear of driving following a motor vehicle accident. CyberPsychology & Behavior 6(3), 329–334 (2003)
13. Wiederhold, B.K., Davis, R., Wiederhold, M.D.: The effects of immersiveness on physiology. In: Riva, G., Wiederhold, B.K., Molinari, E. (eds.) Virtual Environments in Clinical Psychology and Neuroscience, IOS Press, Amsterdam (1998)

14. Wilhelm, F.H., Roth, W.T.: Taking the laboratory to the skies: Ambulatory assessment of self-report, autonomic, and respiratory responses in flying phobia. Psychophysiology 35(5), 596–606 (1998)
15. Wilson, G.F., Russell, C.A.: Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. Human Factors 45(5), 635–643 (2003)
16. Wilson, G.F., Lambert, J.D., Russell, C.A.: Performance enhancement with real-time physiologically controlled adaptive aiding. In: Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting, Santa Monica, vol. 3, pp. 61–64 (1999)
17. Yokota, M., Berglund, L.G., Santee, W.R., Buller, M.J., Hoyt, R.W.: Modeling physiological responses to military scenarios: initial core temperature and downhill work. Aviation, Space, and Environmental Medicine 76(5), 475–480 (2005)

Author Index