This Exhibit is SECRET//SI//NOFORN — Total Funding in Millions

CCP_00083 - Number of voice files rendered searchable by text queries.						
Projects: Analytics Modernization						
Fiscal Year	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**
Funding	-	-	-	$423.8	$512.5	$427.4
Target	1,000,000	300,000	500,000	350,000	350,000	400,000
Result	256,000	263,000	253,871	285,337	-	-
Results and Expected Performance: In FY 2011, the FY 2011 target was not met due to decommissioning of VoiceRT systems deployed outside NSA-Washington. The decommissioning was in preparation for deployment of a more robust voice analytic system in FY 2012. Processing volumes fluctuated during FY 2011. By FY 2013, NSA/CSS expects to have operationalized a more robust voice processing capability based on speech-to-text keyword search and paired dialogue transcription.						
CCP_00103 - Average cost to deliver an analytic hour equivalent for Level 0, Level 1, Level 2, and Level 3 activities (in dollars per hour).						
Projects: Analytic Operational Support, Analytics Modernization						
Fiscal Year	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**
Funding	-	-	-	$595.3	$688.0	$607.0
Target	1.12	1.04	.93	.84	.22	.21
Result	.57	.70	.29	.23	-	-
Results and Expected Performance: In FY 2011, the voice analytic system at NSA/CSS Washington (NSAW) processed 345,000 voice cuts. This growth in machine translation (MT) services utilization resulted in greater than expected analytic efficiency. Due to higher than expected analytic efficiency from the growth in MT services utilization, NSA/CSS is adopting more ambitious future targets for this measure. In FY 2013, more robust voice analytic and machine translation systems will be in operation across the SIGINT enterprise, analyzing greater volumes of data at higher speeds. Productivity increases will further lower the analytic hour equivalent cost.						
CCP_00169 - Percentage of intended storage capacity available as a result of the enterprise distributed database architecture.						
Projects: Analytics Modernization						
Fiscal Year	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**
Funding	-	-	-	$423.8	$512.5	$427.4
Target	1	2	44	75	100	-
Result	0	1	48	55	-	-
Results and Expected Performance: The FY 2011 target was not achieved because a major deployment was rescheduled for late CY 2011 due to a decision to move the system into a government-leased facility as well as a system change that required additional engineering. NSA/CSS expects to complete the deployment of its content storage capacity in FY 2012 allowing NSA/CSS to establish its operational content storage baseline for analyst use. Therefore, this measure will be discontinued, effective FY 2013.						
CCP_00179 - Percent of analysts that have completed digital network intelligence (DNI) core training curriculum.						
Projects: Linguists/Translators						
Fiscal Year	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**
Funding	-	-	-	$228.1	$226.8	$217.5
Target	-	-	Baseline	60	55	55
Result	-	-	42	53	-	-
Results and Expected Performance: The target was not met due to an unexpected rate of analyst attrition toward the end of FY 2011. While NSA/CSS fell short of its target, the number of analysts who received DNI core training increased by 500 each quarter. In FY 2013, NSA/CSS will continue to identify DNI core training gaps and address them with more DNI-specific development plans. This will ensure NSA/CSS analysts are capable of meeting evolving mission needs.						
CCP_00189 - Percent of languages in the civilian NSA Language Reserve Program (LRP), as identified for operational or potential surge requirements, with a capability of three or more analysts.						
Projects: Linguists/Translators						
Fiscal Year	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**
Funding	-	-	-	$228.1	Discontinued	-
Target	-	-	-	34	Discontinued	-
Result	-	-	-	28	Discontinued	-
Results: In FY 2011, the target was not met due to the high operations tempo, which resulted in mission personnel being unavailable to attend scheduled training classes. This measure is discontinued effective FY 2012 and replaced by a new measure, AP_00030.						

THIS PAGE INTENTIONALLY LEFT BLANK

(U) RESEARCH & TECHNOLOGY
(U) HUMAN LANGUAGE TECHNOLOGY RESEARCH

This Exhibit is SECRET//NOFORN									
	FY 2011[1] Actual	FY 2012 Enacted			FY 2013 Request			FY 2012 — FY 2013	
		Base	OCO	Total	Base	OCO	Total	Change	% Change
Funding ($M)	26.4	31.0	3.0	34.0	26.0	3.4	29.4	-4.7	-14
Civilian FTE	8	8	—	8	8	—	8	—	—
Civilian Positions	8	8	—	8	8	—	8	—	—
Military Positions	—	—	—	—	—	—	—	—	—

[1]Includes enacted OCO funding. Totals may not add due to rounding.

(U) Project Description

(U//FOUO) The Human Language Technology (HLT) Research Project provides a coherent, concentrated focus on language analytics to exploit the volume, variety, and velocity of communications that the SIGINT system collects. HLT Research conducts research that supports the goals of the NSA/CSS' Analytic Modernization effort. This Project complements NSA/CSS initiatives to strengthen the language analyst workforce by providing the technologies that serve as force multipliers for analysts.

(U//FOUO) The HLT Research Project has an HLT Center of Excellence (COE) at Johns Hopkins University to promote academic and industry interest in intelligence challenges and attract world-class talent to work on IC HLT problems. The HLT COE focuses on critical intelligence needs that are not adequately addressed by commercial technology or other government programs. The HLT Research Project also leverages programs at the Defense Advanced Research Projects Agency (DARPA) and the Intelligence Advanced Research Projects Activity (IARPA). DARPA and IARPA programs provide foundational HLT capabilities in automatic content extraction, speech-to-text, machine translation, summarization, and question answering. The HLT Research Project conducts research and advanced development necessary to bridge research results from DARPA's and IARPA's efforts to SIGINT applications. This Project includes the Human Language Technology Research Sub-Project.

(U) Base resources in this project are used to:

- (S//SI//REL TO USA, FVEY) Research and develop voice, text, video and image analytics to enable fundamental language exploitation capabilities for all types of communication, regardless of medium.
- (S//SI//REL TO USA, FVEY) Increase the number of languages, accuracy, and speed of results for keyword search from machine-generated transformations of speech-to-text.
- (S//SI//REL TO USA, FVEY) Conduct research and advanced development on automatic document image analysis, particularly for handwritten documents, an extreme technical challenge. The primary emphasis is on core capabilities to enable triage and keyword search on the diverse kinds of documents found in intercept, including language and script identification and handwritten document detection, segmentation, and analysis.
- (U//FOUO) Research analytics that automatically analyze the linguistic content of communications. This area comprises several technologies, including content extraction and machine translation. Content analytics identifies and extracts information from language communications, turning a mass of unstructured text into usable metadata.

- (TS//SI//REL TO USA, FVEY) Research, design, and develop analytics that enable deployment of HLT capabilities nearer to the point of collection within the SIGINT system.
- (U//FOUO) Support collaborative research into human language exploitation and machine learning with commercial and academic partners.
- (U//FOUO) Develop test and training data to support scientific research and evaluation.
- (U//FOUO) Provide and maintain a computer lab to support in-house algorithm development, evaluation, and proof-of-concept demonstrations of promising solutions.
- (U//FOUO) Sustain support activities that foster cross-organizational and cross-discipline collaboration in solving hard technical problems critical to the success of NSA/CSS' SIGINT and cyber missions as well technical health of the workforce.

(U) There are no new activities in this Project in FY 2013.

(U) OCO resources in this project are used to:

- (TS//SI//REL TO USA, FVEY) Enable machine translation research and new speech processing capabilities for Afghanistan and Pakistan dialects using state-of-the-art research findings in less-common languages and by developing new language and dialect models.

(U) The CCP expects this Project to accomplish the following in FY 2013:

- (S//REL TO USA, FVEY) Develop and deploy speech-to-text models for additional languages, where the languages will be selected according to corporate NSA/CSS priorities, language analyst preparation, and scientific assessment of technology readiness. [CCP_0106]
- (S//REL TO USA, FVEY) Extend name-finding solutions to support named-entity extraction for 12 additional languages, to include at least three languages that are less-commonly taught. Create and demonstrate solutions in three to five languages for the much harder problem of extracting relations between entities. These capabilities will yield automated solutions to uncover pertinent facts within both unstructured written communications and spoken communications that have been transformed into text. [CCP_0106]
- (U//FOUO) Design techniques to reduce by 25 percent hand-annotated data required to develop models in support of speech-to-text solutions. [CCP_0106]
- (S//REL TO USA, FVEY) Research, develop, and demonstrate solutions for cross-lingual entity disambiguation to enable analysts to perform language independent retrieval of communications to, from, or about persons of interest from multi-lingual SIGINT data sets. [CCP_0106]

(U) Changes From FY 2012 to FY 2013:

(S//NF) Human Language Technology Research: -$4.7 million (-$5.1 Base, +$0.4 OCO). The aggregate decrease is the result of:

- (U) Increases:
  - (S//NF) $0.4 million in Overseas Contingency Operations (OCO) accelerates new speech processing capabilities and associated analyst applications for Afghanistan and Pakistan dialects.

- (U) Decreases:
  - (S//NF) $5.0 million due to a FY 2012 Congressional add not sustained in FY 2013.
  - (S//NF) $0.1 million due to a planned programmatic reduction in travel and training.

Human Language Technology Research Project Budget Chart FY 2013 Budget Request by Appropriation Account This Exhibit is SECRET//NOFORN			**Funds — Dollars in Millions**		
Subproject	**Description**	**Resourcing**	**FY 2011**	**FY 2012**	**FY 2013**
Operation and Maintenance, Defense-Wide		***Funds***	***—***	***—***	***1.12***
		Positions	***—***	***—***	***8***
Human Language Technology Research	Pay and Benefits	Base	—	—	1.12
		Positions	—	—	8
Research, Development, Test, and Evaluation, Defense-Wide		***Funds***	***26.36***	***34.03***	***28.23***
		Positions	***8***	***8***	***—***
Human Language Technology Research	Communications and Utilities	Base	0.06	0.04	0.04
	Contract Services	Base	24.35	28.07	23.36
		OCO	—	3.00	3.40
	Equipment	Base	0.57	1.76	1.36
	Pay and Benefits	Base	1.20	1.09	—
	Travel and Transportation	Base	0.17	0.07	0.07
		Positions	8	8	—
					Totals may not add due to rounding.

THIS PAGE INTENTIONALLY LEFT BLANK

NATIONAL SECURITY AGENCY
CENTRAL SECURITY SERVICE

(U) Classification Guide for Human Language Technology (HLT) Models 2-20

Effective Date: 18 May 2011

CLASSIFIED BY: [redacted]
Deputy Director for Analysis and Production

Classification Category: 1.4 (c)

Declassify On: 25 years*

ENDORSED BY: [redacted]
Deputy Associate Director for Policy and Records

CLASSIFICATION GUIDE TITLE/NUMBER:
(U) Human Language Technology (HLT) Models, 2-20

PUBLICATION DATE: 18 May 2011

OFFICE OF ORIGIN: (U) R66E, Human Language Technology Research

POC: (U//FOUO) [redacted] R66E; 961-3032s

ORIGINAL CLASSIFICATION AUTHORITY: (U//FOUO) [redacted] Deputy Director for Analysis and Production

Description of Information	Classification/ Markings	Category	Declass	Remarks
A. (U) General				
A.1. (U) The fact that NSA/CSS has created HLT models used for: • Gender recognition • Language • Language variety/dialect recognition • Speaker recognition • Speech-to-text processing • Speech activity detection • Anomaly detection • Phonetic recognition	UNCLASSIFIED	N/A	N/A	
A.2. (U) The fact that HLT models are obtained, at least in part, by aggregating statistics derived from SIGINT collection	UNCLASSIFIED	N/A	N/A	
A.3. (U) The fact that HLT models allow for collected audio files to be sorted and prioritized for linguists	UNCLASSIFIED	N/A	N/A	
A.4. (U) The fact that statistics in a model can be generated from one or many audio files	UNCLASSIFIED	N/A	N/A	
A.5. (U) The fact that new models are regularly generated, adding to the aggregate nature of the model	UNCLASSIFIED	N/A	N/A	
A.6. (U) The fact that SIGINT voice collection (not further identified) can be identified as: • male or female • a specific language • a specific language variety/dialect • a specific speaker • a sequence of words • speech or nonspeech	UNCLASSIFIED	N/A	N/A	(U) Further details such as which specific language, or dialect, or speaker are classified. Consult applicable SIGINT guidance.
A.7. (U) HLT models used for: • Gender recognition • Language recognition	See Remarks			(U//FOUO) The classification of HLT models used for Gender and Language Recognition is dependent upon the classification of the messages used to train the model, up to SECRET//REL TO

				USA, AUS, CAN, GBR, NZL. Although it is possible that the messages used to train the model may have a higher classification and/or more restrictive releasability than SECRET//REL, the original audio cannot be recovered from the model. SECRET//REL is sufficient to protect this type of model. (U) The Deputy Director for Analysis and Production may approve, on a case-by-case basis, foreign release of models containing otherwise non-releasable information.
A.8. (U) HLT speaker recognition models	See Remarks.			(U) Consult applicable SIGINT guidance: Classification and foreign releasability should be in accordance with the highest classification and most restrictive releasability that applies to the targeted entities used in the model. (U) The Deputy Director for Analysis and Production may approve, on a case-by-case basis, foreign release of models containing otherwise non-releasable information.
A.9. (U) HLT acoustic models used for • Speech-to-text • Phonetic tokenization	See Remarks.			(U//FOUO) The classification of HLT acoustic models is dependent upon the classification of the messages used to train the model, up to SECRET//REL TO USA, AUS, CAN, GBR, NZL. Although it is possible that the messages used to train the model may have a higher classification and/or more restrictive releasability than SECRET//REL, the original audio cannot be recovered from the model. SECRET//REL is sufficient to protect this type of model. (U) The Deputy Director for Analysis and Production may approve, on a case-by-case basis, foreign release of models containing otherwise non-releasable information.
A.10. (U) HLT language models used for • Speech-to-text	See Remarks.			(U) Consult applicable SIGINT guidance: Classification and foreign releasability should be in

• Phonetic tokenization				accordance with the highest classification and most restrictive releasability that applies to the targeted entities and/or content of the messages used in the model. (U) The Deputy Director for Analysis and Production may approve, on a case-by-case basis, foreign release of models containing otherwise non-releasable information.
A.11. (U) Speech activity detection models using syllable rate speech activity detection (SRSAD)	UNCLASSIFIED// FOR OFFICIAL USE ONLY	N/A	N/A	
A.12. (U) Anomaly detection models	UNCLASSIFIED// FOR OFFICIAL USE ONLY	N/A	N/A	
B. (U) Model Output				
B.1 (U) Output of language recognition models	UNCLASSIFIED// FOR OFFICIAL USE ONLY	N/A	N/A	(U) Results generally indicate the recognized language and the degree of confidence in the determination, e.g. "Farsi with 90% confidence." This information may require protection as classified when combined with other details regarding the input data.
B.2. (U) Output of gender recognition models	UNCLASSIFIED// FOR OFFICIAL USE ONLY	N/A	N/A	(U) Results generally indicate the recognized gender and the degree of confidence in the determination, e.g. "Male with 75% confidence." This information may require protection as classified when combined with other details regarding the input data.
B.3. (U) Output of speaker recognition models	See Remarks.			(U) Classification and foreign releasability of the results should be the same as the input data.
B.4. (U) Output of acoustic speech-to-text and phonetic tokenization models	See Remarks.			(U) Classification and foreign releasability of the results should be the same as the input data.
B.5. (U) Output of language speech-to-text and phonetic tokenization models	See Remarks.			(U) Classification and foreign releasability of the results should be the same as the input data unless the results reveal specific information used in the model that is protected at a higher level than the input data; in this case, the results require protection at the level of the model.

(U) Note: Declassification in 25 years indicates that the information is classified for 25 years from the date a document is created or 25 years from the date of this original classification decision, whichever is later.

RT[10] Overview

June 2006

Current Challenges – Why Innovate?

- **Network is converging and dataflow is increasing**
- Pushing everything back means front-end filtering → dropped data
- Substantial lag to process, store, and query data
- Restricted geo-spatial capabilities (not just because of the hardware)
- Manual correlation between SIGINT, HUMINT, SIGACTs
- Non-integrated toolsets
- No comprehensive theater knowledge base
- Non-optimal collaboration between analysts
- Manually intensive production processes
- Cannot scale / work targets in volume
- Reaching limits of legacy systems
- Analysis takes a lot of time

Think: Atlantic Monthly vs. CNN scrolling bar

RT-10 Goals

- Overall Objective: Order of magnitude improvements in real-time SIGINT architecture for the U.S. Cryptologic System, initially focused on national and tactical intelligence in Baghdad, to enable **better** decisions in **less** time

- Providing:
  - Access to **more comprehensive** data
  - Immediate access to local data sets, with query back to NSAW
  - Integrated Analytic Workflow, with **better** tools
  - Real time Alerting: National and Tactical
  - Automation of tasks → Query to Dissemination
  - Distributed Analytic Collaboration
  - Scalability
  - **Integration** across brigade-level SIGINT capabilities

Tools and Workflow

- **Better Tools**
  - New relationship visualization w/temporal capabilities
  - Real-time geo-spatial alerting framework
  - Web-based applications optimized for speed in a distributed environment
  - Partnership with Green Dragon to identify and inject new COTS/GOTS technologies in **much less** time
- **Integrated Framework**
  - Work in any tool of choice, seamlessly switch to alternative views (think development → presentation)
  - Automated, one-click mentality from query to dissemination
  - Developers available to react to analyst needs and inject new capabilities

Substantial Improvements in Data Access

- **Initial**
  - Traditional Data Sources (Global Reach touching NSA databases)
  - SCS GSM collection
  - Tactical GSM Accesses
  - Checkpoint Data
  - HUMINT / All-Source derived SIGINT Selectors (parsed CIA TDs, DOD IIRs, TAREX, DOCEX)
  - Local knowledge base
- **Future**
  - Fully-integrated Iraqi DNI Data flows (initially accessible through separate web interface)
  - OBELISK / LETC GSM Coverage
  - WISPYKNIT, VICTORYUNIFORM and other special source

Think: Know everything we collectively know, and faster

VoiceRT: Index / Search of Voice Cuts

- Goal: Better filtering and selection using latest generation of voice-processing technologies
  - Perform phonetic indexing on 1 million voice cuts per day
  - Run incoming cuts against 1000 individual voice prints to drive real-time filtering and selection
  - Optimizes linguist scarce time, does not replace linguists

- Increases efficiency of available linguists
- Allows analysts to affect collection priorities and react to changing linguistic / word patterns
- Possible future integration with checkpoint collection (voice / biometrics)

Real time Alerting: National and Tactical

- **Real-time alerting on hard selectors**
  - Creating a knowledge base within the collection architecture
  - Drives selection and filtering
  - Provides relevant information to war fighter in seconds
- **Algorithms to Detect and Alert from Patterns of Interest**
  - Constant enrichment of incoming data flows based on NSA and GCHQ-developed algorithms
  - Robust framework to allow analysts to submit / modify / reject existing techniques
  - Capability to extend algorithms to correlate and react to friendly actions, geospatial and geotemporal vicinity, etc.

Dedicated effort to identify and detect new potential targets based on known behavioral patterns

Automation of Standard Tasks

- **One-Click Report Generation**
  - Chaining diagram
  - Products containing target
  - Frequent calling list
  - Temporal trends
  - Geospatial trends

- **One-Click Alert-to-Analysis**
  - Alerting framework fully integrated with analytical toolsets
  - Geospatial / Temporal / Network views of data

- **One-Click Analyst Actions**
  - Drive collection through interface to EDB / Keycard
  - Effortlessly affect knowledge base confidence / details

Checkpoints

- Provide advanced sensors to generate checkpoint metadata:
  - Active cell phone interrogation
  - Active RF Illumination: goal to fingerprint vehicles, identify threats (artillery shells, ammunition, gun barrels, electronic triggers)
  - 360 degree imagery
  - Chemical and radiological detectors
- Fed real time to tip and cue other Ints

- Proof of concept in vicinity of BIAP, tentatively checkpoint 538 on Route Irish
- Operational test, tentatively checkpoint 502 near Abu Ghurayb

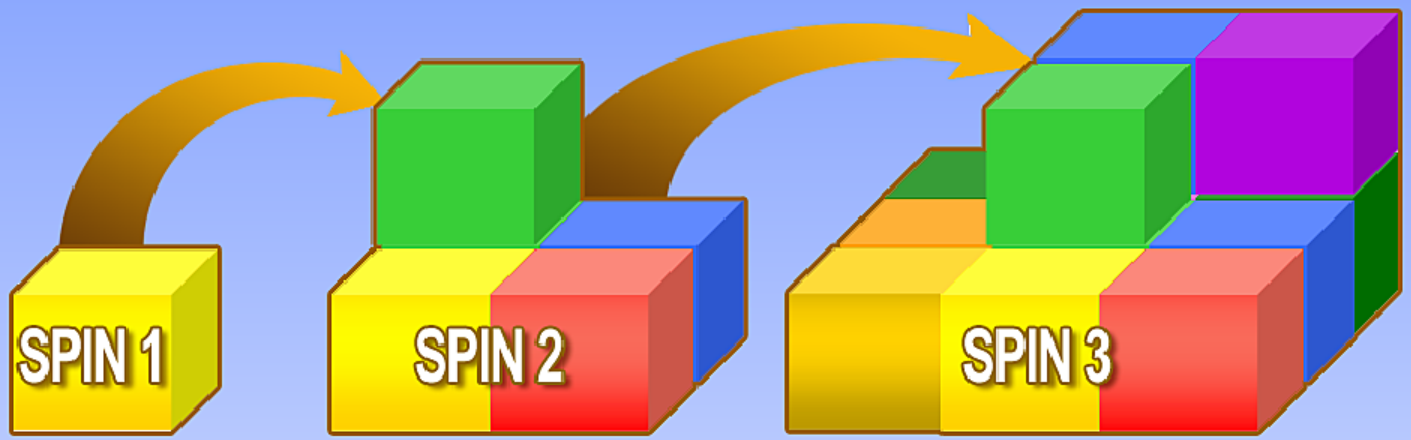
Implementation

- **Construct the JIOC-I "SIGINT Brain"**
  - Distributed Databases in Baghdad and Ft Meade
  - Aggregate Metadata from tactical and national collection, focusing on GSM for initial efforts
  - Massive data flows: 50 Million+ GSM metadata events / day
  - Content access from all possible collectors
  - Integration: "Know what we know"

- **Timeline**
  - End June – Initial site surveys, theater coordination
  - 15 July – Hardware Ships
  - 1 August – Hardware Arrives, People begin arrving
  - 15 August – System Online
  - 31 August – Data Flowing
  - 15 September – IOC

Achieving Success with Spins

- **Spin Methodology:**
  - Iterative activity consisting of a series of spins
  - Each 90-day spin expands capability
  - Demonstration of integrated capabilities
  - Application of new and existing technologies
  - Make discoveries and apply lessons learned to future spins



RT[10] Spin Schedule

Spin 1: August 2006

- System Installed
- JUGGERNAUT Data
- Initial Software Testing

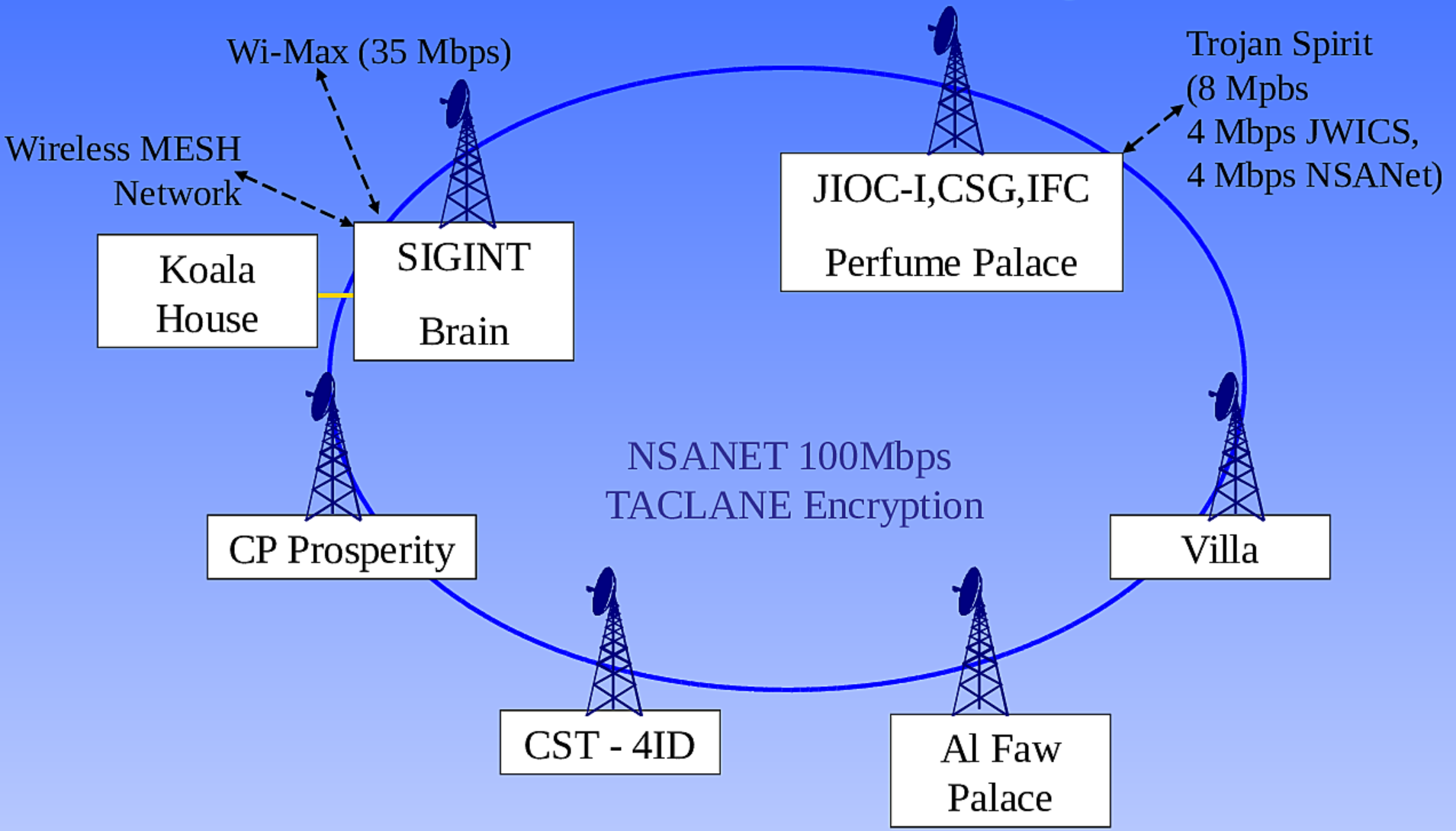
Spin 2: Nov 2006

- Demonstration of integrated capabilities
- Cable / FORNSAT Integration
- Enhance Checkpoint Capability
- Analyst-Identified Areas of Improvement

Spin 3: Jan 2007

- Analyst-Driven Modifications
- Next-Generation Analytical Tools

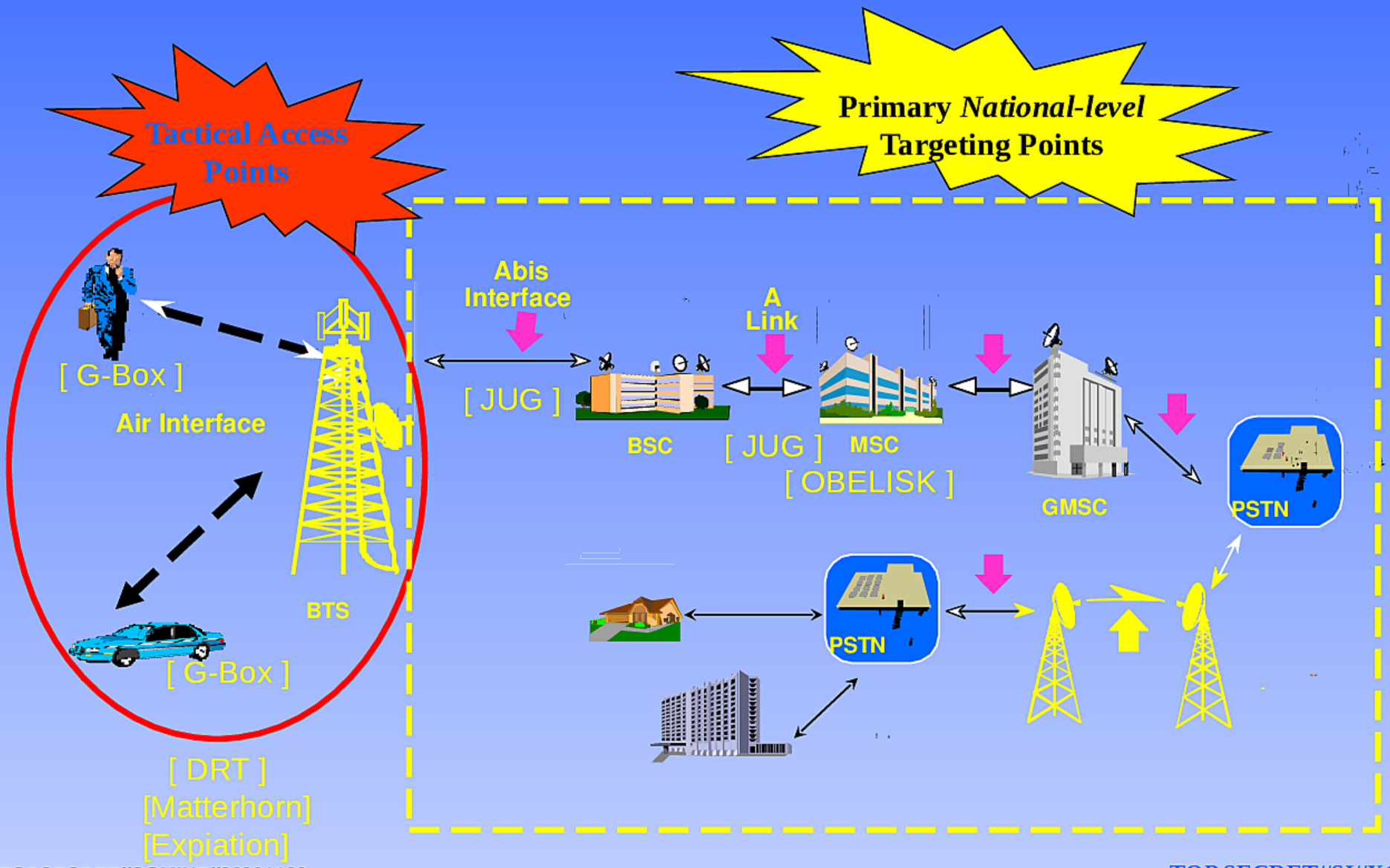
RT[10] Line of Sight Microwave Network
Wi-Max, Mesh, NSAnet Connectivity



RT[10] Analytic Nodes

- Iraq
- MOC
- NSA Product Lines
- NSA-G
- COBRA FOCUS

GSM Architecture & Access Points



VoiceRT: Index / Search of Voice Cuts

- Goal: Better filtering and selection using latest generation of voice-processing technologies
  - Perform phonetic indexing on 1 million voice cuts per day
  - Run incoming cuts against 1000 individual voice prints to drive real-time filtering and selection
  - Optimizes linguist scarce time, does not replace linguists
- Increases efficiency of available linguists
- Allows analysts to affect collection priorities and react to changing linguistic / word patterns
- Possible future integration with checkpoint collection (voice / biometrics)

(U) For Media Mining, the Future Is Now! (conclusion)

FROM: [redacted] and
Human Language Technology (S23)
Run Date: 08/07/2006

(S//SI) Media Mining Across a Wide Range of Languages

(S//SI) One of the challenges in deploying this Media Mining HLT is the need to cover the very broad range of languages. Unfortunately, most of the languages of interest to the Agency are not of interest to commercial concerns because they are not likely to be profitable, and businesses run on profit.

(S//SI) Though COTS products such as NEXminer have covered commonly-taught, "dense" languages such as English and Spanish, and have made great inroads lately into a few less-commonly-taught languages and dialects found in the Middle East, it is unclear that any COTS product will ever cover the vast inventory of languages that NSA analysts are required to understand. Therefore, the HLT PMO is developing an enhancement of this Media Mining technology that can process over 90 languages using a combination of language-specific and universal phones. This agency capability, developed within R64, the Human Language Technology Research Group, is known as Universal Phonetic Recognition (UPR).

(S//SI) New languages can be easily added to the technology by drawing on Agency linguistic knowledge of a language combined with publicly available language resources. As world events shape our language needs, UPR provides a way to respond within minutes to new language needs, for example to support the GWOT.

(U) IVE: Technology that Can Separate the Wheat from the Chaff

(S//SI) A second, equally important enhancement under development is the ability for this HLT capability to predict what intercepted data might be of interest to analysts based on the analysts' past behavior. Much like the way in which popular sites like amazon.com are able to track and predict buyer preferences, integration of Intelligence Value Estimation (IVE) on both SRI and message content, offers the promise of presenting analysts with highly enriched sorting of their traffic. Imagine if you came to work each day knowing that the best five intercepts needing transcription were sitting at the top of your queue waiting for you.

(S//SI) Of course, such Media Mining IVE capabilities need not be limited to SRI and key word searches. In collaboration with S202B, Analytic Technologies for the Enterprise, the HLT PMO Media Mining team is also developing new metadata analysis capabilities based on language, speaker, gender, and dialect identification, presenting this information to analysts through conventional query tools such as UIS. Advanced programs like RT-10 are integrating other forms of information, such as geospatial coordinates. RT-10 will also send automatic alerts to analysts when incoming intercept meets certain search criteria.

(S//SI) Voice$_{RT}$ will soon be integrated with standard Agency voice tools such as UIS and HOTZONE. Analysts will be able to configure the tool via the web, and access scores on their traffic using NUCLEON.

(U) Bringing it All Together

(S//SI) The integration of these technologies into an automated system will bring two major innovations: faster response time and improved productivity. Our challenge goal is to "index, tag, and graph" all incoming intercept, and this will soon be within reach. Using HLT services, a single analyst will be able to sort through millions of cuts per day and focus on only the small percentage that is relevant. The amount of collection can be increased orders of magnitude without further stressing the analyst population, allowing the Agency to cast a much wider SIGINT net and taking in a much richer catch.

(S//SI) And again, the power of HLT is truly realized through integration of multiple SIGINT technologies. In the future, we will further develop technologies such as word search to support cross-lingual queries. Sites that lack expertise in a given language will be able to issue queries in English and receive results translated from the target language back into English. This marriage of word search and Machine Translation has great potential as a force multiplier. Mapping meaning and tradecraft across languages will be a key challenge here.

(S//SI) Similarly, because a search term will be tagged with a "semantic class identifier," such as "place name," it will be relatively straightforward to integrate this technology with the Enterprise Knowledge System (EKS) and allow sophisticated capabilities such as social network analysis to operate on voice content. In the HLT PMO long-term vision, analysts will be able to construct complex queries, such as, "Where is the mayor of Baghdad?" or "Show me all the intercept containing information about explosive devices that occurred yesterday in the downtown area of Baghdad near the Al-Rashid Hotel," and obtain answers directly in English, or in their foreign language if they prefer, with a link to the documents containing the answers.

(U//FOUO) **We are entering a golden age for HLT.** Powerful and inexpensive computers, high-speed networking, and advanced algorithms are being combined to revolutionize the analyst desktop.

(U//FOUO) For more information about these capabilities, please contact the HLT PMO office ("go HLT" or call (s)[redacted]) .

(U) For Media Mining, the Future Is Now!

FROM: [redacted] and
Human Language Technology (S23)
Run Date: 08/01/2006

(TS//SI) In the first article on the Human Language Technology Program Management Office's (HLT PMO) activities and plans, we explained that we have five Strategic Thrusts. In this article, we will focus on the most active and fast-paced of the five: **Media Mining.** Its goal is to provide seamless access to information no matter what the information's source may be -- audio, image, or text. Right now over two hundred analysts have access to some Media Mining capabilities.

(S//SI) Near-Real-Time Alerts: RT-10

(S//SI) Integration of diverse information sources to produce near-real-time alerts is a major goal of a new Agency-wide program, **RT-10.** RT means REAL TIME, and 10 refers to reducing the time between collection and the generation of actionable intelligence an order of magnitude in each spin of the project.

(S//SI) The first deployment of RT-10 to the JIOC-I in Baghdad in 4th quarter 2006 will focus on integration of diverse information sources, including GSM voice intercept and geospatial coordinates, to reduce the time required to generate actionable intelligence.

(S//SI) New Voice-Services Platform: Voice_{RT}

(S//SI) The HLT PMO is collaborating with RT-10 on the development of a new voice services platform, **Voice_{RT}**. The first deployment of Voice_{RT}, which is architecturally-based on an Army INSCOM* prototype known as ALICAT, will be operational in the Baghdad node of RT-10 in September 2006. This system is designed to index and tag 1 million cuts per day, and provide auxiliary HLT services such as language, dialect and speaker identification. The combination of these technologies with other RT-10 capabilities, such as geospatial coordinates, will provide a unique ability to generate actionable intelligence quickly and accurately.

(S//SI) Voice_{RT} is a tool that allows analysts to perform keyword searching on voice content.

(S//SI) Voice Word-Search Capabilities

(TS//SI) The HLT PMO's Media Mining Thrust began as an effort to bring word-search capabilities (e.g., "Google for Voice") to Voice Language Analysts to make it easy for them to locate intercept rich in intelligence data. Voice word search technology allows analysts to find and prioritize intercept based on its intelligence content in much the same way as they now search text in PINWALE. For example, in the Global War on Terrorism (GWOT), analysts can locate intercept dealing with explosive devices by searching for common terms such as "*operation*" or "*detonator*," as well as more subtle terms about materials ("*hydrogen peroxide*"), place names ("*Baghdad*"), or people ("*Musharaf*").

(S//SI) The first generation of this technology has been centered around Commercial-off-the-Shelf (COTS) software, **NEXminer**, developed by a startup company, Nexidia. The system is designed

to support both **real-time searches**, in which incoming data is automatically searched by a designated set of dictionaries, and **retrospective searches**, in which analysts can repeatedly search over months of past traffic. The former capability allows the tool to function as a near real-time tipper. The latter capability allows analysts to rediscover important intelligence information and to refine their search strategies. This can be especially important in cases where pieces of a SIGINT "puzzle" become apparent and an analyst needs to go back to previous messages to see if other unnoticed pieces can be found.

(S//SI) This tool is very effective because it integrates high-performance speech processing technology with a most important agency resource, analyst knowledge of targets and missions. This technology was initially introduced to the analyst community in 2004 as a prototype, RHINEHART, which had been developed by SIGDEV Strategy and Governance (SSG).

(S//SI) RHINEHART now operates across a wide variety of missions and languages, and is used throughout the NSA/CSS Enterprise. One recent example of RHINEHART success occurred when Persian GWOT analysts searched for the words "negotiations" or "America" in their traffic, and RHINEHART located a very important call that was transcribed verbatim providing information on an important Iranian target's discussion of the formation of the new Iraqi government.

*Notes: (U) INSCOM = US Army Intelligence and Security Command

(U) Watch for the conclusion of this look at media mining, coming soon...

(S//SI//REL) How Is Human Language Technology (HLT) Progressing?

FROM: (U//FOUO) [redacted]
Language Analysis Modernization Lead (S2)
Run Date: 09/06/2011

(S//SI//REL) Editor's intro: At the SID town hall meeting of February 2011, [redacted] (pictured) briefed on Human Language Technology, i.e., tools that sort through SIGINT voice collection and automatically find the most promising nuggets, thereby saving linguists countless hours. What's happened with HLT since that time?

(S//SI//REL) In 2011 we deployed HLT Labs to Afghanistan, NSA Georgia, Latin American SCS sites, and NSA Texas.

(U) **Afghanistan-area targets**

(S//SI//REL) Afghan Regional Operating Cryptologic Center (AROCC) analysts started using HLT Labs to track their targets in April, and when the analytics were successfully used to find new information, the mission was expanded to include international teams.* The Afghanistan deployment boasts some technological firsts associated with cloud computing** and includes the full suite of analytics with Pashto speech-to-text (STT). Recently French analysts in the ARC were able to find target speakers on new selectors using speaker recognition.

(S//SI//REL) Our deployment to NSA Georgia enables us to partner with analysts to assess the performance of our newest STT models: Pashto and Farsi. These languages have limited training data which creates challenges for STT, and we have been focused on finding applications that are beneficial even for these low-resource languages. NSA-Georgia traffic includes noisy VHF collections which seriously degrade analytic performance; however, analysts can still find target speaker cuts on unknown frequencies.

(U) **Spanish-speaking targets**

(S//SI//REL) Spanish is the most mature of our speech-to-text analytics, and has higher keyword-search accuracy than other deployed STT models. We've had great success searching for Spanish keywords at NSA Texas and Latin America SCS sites.

(S//SI//REL) For example, in early August a new NSA Texas user applied keyword search the morning after his training to find a previously unreported cut from a drug trafficking target. Likewise, the OIC of one of the Latin American SCS sites recently reported he was able to find foreign intelligence regarding a Cuban official in a fraction of the usual time. His comment: This same example could be used over and over by many that have to go over countless voice cuts to finally dig that gold nugget that will turn into a report.

(U) **Development work continues**

(U//FOUO) The R6 research team is working to add new applications, improve keyword search capability, enhance analytics, add new languages, and refine the user interface. Recently the Summer Camp for Applied Language Exploration (SCALE) -- a joint NSA Johns Hopkins University exercise -- investigated new ways to use the results of HLT analytics from existing targets to find new targets. Research is also working closely with the SPIRITFIRE (voice analytics)

and TransX (translation, transcription and transliteration) efforts to ensure HLT Labs capabilities are included in the corporate solution for enterprise deployment in 2012.

(U//FOUO) More information about HLT Labs is available here.

(U//FOUO) See a related SID*today* article about HLT here.

(U) Notes:

* (S//REL) The international teams were from the Analysis and Research Cell (ARC), Task Force 310, and Combined Joint Special Operations Task Force (CJSOTF).

** (S//SI//REL) Specifically, the Afghan deployment is the first use of DISTILLERY and CLOUDBASE on a GHOSTMACHINE platform.

(U//FOUO) Coming Soon! A Tool that Enables Non-Linguists to Analyze Foreign-TV News Programs

FROM: [redacted]
Center for Time-Sensitive Information (S2413)
Run Date: 10/23/2008

(U//FOUO) Have you ever wanted to use foreign-TV news broadcasts to enhance your SIGINT, but couldn't because you didn't understand the language? Soon you can! It is currently only available to NSOC Desk Officers until logistical issues are resolved. However, the goal is to migrate Enhanced Video Text and Audio Processing (eViTAP) into other areas. Additional information regarding how to get accounts will be announced in the coming year, so stay tuned.

(U//FOUO) EViTAP is a fully-automated news monitoring tool. The key feature of this Intelink-SBU-hosted tool is that it analyzes news in six languages, including Arabic, Mandarin Chinese, Russian, Spanish, English, and Farsi/Persian. "How does it work?" you may ask. It integrates Automatic Speech Recognition (ASR) which provides transcripts of the spoken audio. Next, machine translation of the ASR transcript translates the native language transcript to English. Voila! Technology is amazing.

(U//FOUO) Figure 1: Example of video, native language transcript and English translation

(U//FOUO) This all sounds wonderful, but is it easy to use? Absolutely! EViTAP has an intuitive and easy to use browser-based interface. The User Guide includes everything you need to successfully use the tool. It is perfect for the analyst who usually prefers classroom over manual instruction. The interface provides advanced search and retrieval capabilities, real-time fully automated alerting, ability to create clips from videos, ability to edit transcripts and translations, ability to export video and transcripts to PowerPoint, XML and text formats, and much more.

(U//FOUO) EViTAP's capabilities are far reaching, and go beyond the scope of this article, but it is clear that this tool can significantly enhance SIGINT analysis and reporting. Open Source is becoming more significant in the Intelligence Community, and eViTAP is an open source resource that can play a big role in enabling SIGINT prosecution.

(U//FOUO) Want to learn more about eViTAP or other open source resources? Simply type "go AIRS" in your browser window and explore the information there. AIRS, Advanced Intelligence Research Services, provides a multitude of open source and all collateral products and services. EViTAP is a new open source tool AIRS introduced at NSA.

(U//FOUO) POC: [redacted]

(S//SI) Dealing With a 'Tsunami' of Intercept

FROM: [redacted]
Human Language Technology (S23)
Run Date: 08/29/2006

(S//SI) Everyone knows that analysts have been drowning in a tsunami of intercept whose volume, velocity and variety can be overwhelming. But the Human Language Technology Program Management Office (HLT PMO) can predict that in the very near future **the speed and volume of SIGINT will increase even more, almost beyond imagination.** And we are working on ways to help analysts deal with it all.

(S//SI) Of the HLT PMO's five Strategic Thrusts, the one that addresses this problem is **High Speed/ High Volume.** It must deal with today's collection and must plan for tomorrow's. The current collection environment is characterized by huge amounts of data, coupled with severely limited capability to send material forward, and extremely limited number of queries that exactly describe messages of value. That means we are capable of finding huge amounts of data, much of which is not what we really want, and that we cannot send it all back for analyst processing.

(TS//SI) To plan for tomorrow, High Speed/ High Volume is in line with changes in the overall NSA/CSS systems, particularly TURBULENCE and TURMOIL because when they become a reality in the near future, we can expect collection capabilities to increase significantly. TURBULENCE is an umbrella cover term describing the next generation mission environment that will create a unified system. TURMOIL is a passive filtering and collection effort on high-speed networks. This is designed to be flexible and can be modified quickly to deliver data in analyst-ready form.

(S//SI) One of High Speed/ High Volume's first efforts is in developing and implementing ways to **push HLT capabilities very close to the collection points of the SIGINT system.** In particular, HLT is about to demonstrate an operational prototype of language identification for Special Source Operations (SSO) Counterterrorism text targets running at line speeds (STM-16) at the packet-level. Resources permitting, HLT analytic processors will automatically generate content-based events for TURMOIL based on language.

(S//SI) HLT processors will demonstrate the ability to characterize very high speed channels based on content, thus enabling analysts to task the SIGINT system to send back messages based on information found in message content, not just on externals. (Externals can be Signal Related Information (SRI) that comes with each message, such as channel, Time Up/Time Down, etc.) Using HLT services, analysts will be able to build more precise descriptions of the data they want. In addition, content-based metadata will allow SIGDEV analysts to run more detailed surveys. HLT services that work on data content at the collection point can also provide indications or warnings that the SIGINT system must adapt its collection strategy.

(S//SI//REL) Resources permitting, High Speed/ High Volume will deploy **capabilities for voice, text, and image data,** and will take advantage of research being done by a number of organizations including the Research Directorate's Coping With Information Overload Office (R6), Disruptive Technologies Office (DTO), and SID/ Analysis and Production's Advanced Analysis Laboratory (AAL). HLT research and transfer of its technology into operations means the development of algorithms that can incorporate HLT capabilities for the processing of elements such as email attachments and VOIP.

(S//SI/REL) The research and technology transfer also may provide "stealthy," low-profile in-target implants for Tailored Access Operations (TAO) or technologies to enable high speed processing in

very low size, weight and power applications for other CLANSIG customers. And, to help address the "unknown unknown" target analysis problem, HLT is investigating techniques and technologies for high volume voice processing so that all voice data can be scanned for key words before it is selected based on phone numbers.

(S//SI) Ultimately, HLT's High Speed/ High Volume will give the analyst greater ability to influence collection and processing much farther forward in the SIGINT system, as well as help the SIGINT system achieve greater overall filtering and selection effectiveness. That means **more analysts wil be getting better SIGINT** at a time when volume and velocity are maximum.

B/7655BA/1400/00006/018/0
7 December 2009

SIRDCC Speech Technology WG assessment of current STT technology

Security Service have asked the SIRDCC Speech Technology Working Group to give its technical assessment of the current state of the art in Speech to Text technology, and how it is likely to develop.

Executive summary

The SIRDCC Speech Technology Working Group has evidence that current state of the art STT technology is capable of providing some business benefit in very specific circumstances. It has still to prove itself in larger-scale applications, but the potential for major benefits in productivity in the future is clear, given sufficient investment in further developing the systems for our target speech.

The Working Group believes that the most effective way to achieve these benefits is to continue to fund research and development activities. Where practical this should be supplemented with small-scale pilot deployments to explore the areas where most immediate business benefit can be got, so as to help focus the R&D investment.

The underlying technology used by all existing state-of-the-art systems is similar, and thus each is in principle capable of obtaining similar results in any given application, given sufficient effort in bespoke development and tuning. However the BBN system currently deployed at GCHQ for the last 5 years and at NSA for longer has proved itself stable, currently outperforms others on the standard measure of word error rate and is therefore recommended for operational pilots in the near term.

The decision as to when and how it is appropriate to deploy an operational pilot in any agency must depend on business decisions internal to that agency, but it is important that we share and collaborate to the fullest extent to minimise costs and maximise benefits.

Context

Security Service and GCHQ have been collaborating on research and development of capability for Speech to Text (STT), also known as Automatic Speech Recognition (ASR), for a number of years under the auspices of the SIRDCC Speech Technology Working Group. The aims are to assess the applicability of the technology to gain business benefit, and to conduct appropriate research and development to advance the technology where needed.

The other members of the Speech WG have a strong interest in the outcome as a means of informing their own future investment decisions.

UK SECRET STRAP1

DARPA evaluation programme

The DARPA evaluation programme, with significant steer from NSA, has been the main driving force behind technology improvements in the field. Unfortunately the results of the evaluations are not put in the public domain, making reference difficult.

Most of the large corpora of transcribed speech were produced under this programme for evaluation purposes: they are made of up rather artificial conversations between speakers (often college students) who are paid to take part.

Cambridge University and BBN have participated throughout the lifetime of the programme: they have joined forces for the current phase (GALE). Both have always been at the forefront. So were Dragon until their collapse and IBM until they pulled out a few years ago. IBM have subsequently re-entered with the stated objective of obtaining better than human performance, and they marginally outperformed the BBN/Cambridge entry in the most recent evaluation.

Other research labs and universities have also taken part but have never done as well as the organisations mentioned above. SAIL have never participated.

The systems used in these evaluations are research software, and not written for use by anyone other than the originating labs. A version of the BBN system is the only exception to this, having been in use at NSA for about 10 years. In this period a lot of effort has been put into giving it at least some robustness and usability, and into making it user-trainable.

Cambridge University have always taken the view that their software was for running on their own site only, though a modular toolkit HTK is publicly available.

To the best of our knowledge Security Service's purchase of Attila from IBM is the first instance of it being trained other than at its originating site, though we have reports that DSTO and CIA are also investigating its performance.

NSA programme

NSA have had the BBN speech-to-text system Byblos running at Fort Meade for at least 10 years. (Initially they also had Dragon.) During this period they have invested heavily in producing their own corpora of transcribed Sigint in both American English and an increasing range of other languages. Their application of English is to COMSEC monitoring. One of GCHQ's hopes is that NSA will give it access to the models being trained on SIGINT data, since NSA have considerable difficulty in releasing the intercept itself. This is one of the motives for GCHQ's adopting Byblos, since models trained by one system cannot be used by another.

GCHQ/Security Service approach

We have pursued our aims in this field in two main ways, evaluating systems as delivered and obtaining training data to seek to improve them. Our goals have been: (1) to evaluate the technology itself and its business applicability; (2) to perform a comparative evaluation of competing systems to decide where best to concentrate our resources.

- **Systems evaluation**

GCHQ has licensed the Byblos system from BBN Technologies, Boston, since 2002. This system was chosen partly because it was the best-performing system in external trials run by DARPA, but most importantly because it was already in use as a research system within NSA, who were also funding much of its development. GCHQ also funded some specific development by BBN in 2006 in order to make it more easily deployable on our systems.

Security Service (C3T) has investigated the performance of speech recognition from IBM. The initial judgement of IBM, made in 2001, was that their technology was not yet ready [1], but their comparative success in DARPA trials in 2004 led to renewed interest from Security Service who arranged for further trials on UK-accented speech by IBM. In 2009 Security Service licensed the IBM Attila system and funded IBM effort to help build and evaluate a speech recogniser specifically for Security Service product.

Security Service (A2K), with funding assistance from GCHQ, has investigated the performance of speech recognition from a European company, SAIL labs of Vienna. SAIL have licensed their system to Security Service and built a speech recogniser for evaluation.

- **Bulk transcription**

It has been recognised for several years that the main obstacle to effective STT of intercepted speech was the mismatch between the models of speech used in STT systems and the intercept. To address this using current STT technology, tens or hundreds of hours of speech must be carefully transcribed at great cost in order to provide training data. There are two deficiencies in current STT systems. Firstly their models of conversational English speech are biased strongly towards US English. Secondly, the material is gathered openly and is not representative of the speech of the majority of our targets.

GCHQ and Security Service have collaborated to acquire, transcribe and share data sets. Most of these have been UK English of various regional accents, obtained commercially, but we also have a substantial corpus of regional Arabic. A small amount (75 hours in total) has been transcribed from intercept. Of this, there is one

significant UK-regional corpus, NIRAD, which is 56 hours of mostly Northern Irish accented speech.

The very high cost of transcription for STT purposes (of the order of £1500 per hour of speech) makes it vital that we continue to collaborate and share as much as possible.

Status in December 2009

- **Systems evaluation**

The NIRAD corpus has been used to train and evaluate all three systems. The results are reported in a joint GCHQ-Security Service paper [2].

The overall figures on word error rate were: BBN 63%, IBM 82%, SAIL 101%. The figures for word accuracy were: BBN 42%, IBM 32%, SAIL 20%. Note that error rate and accuracy do not necessarily add up to 100% as the error rates are normalised with respect to the true transcript and there may be additional words incorrectly inserted by the recogniser.

The analysis shows that the BBN recogniser is better than the IBM recogniser at transcribing words by a significant margin, as measured by the number of words in each speech file that it got correct (better in 58 out of 59 files).

The analysis also shows that by this measure the IBM recogniser is better than the SAIL recogniser by a significant margin (better in 57 out of 59 files).

There is substantial variation in the recognition rates of individual words. See the Appendix for a representative sample of text as transcribed by the BBN Byblos system, and how bespoke training improves the recognition. There is also a table of the best recognised words, other than those which are recognised 100% which are mostly singletons perhaps well-recognised by accident.

For these experiments Byblos was trained by GCHQ staff with no BBN involvement. The SAIL system was trained by its developers. Attila was trained by Security Service with assistance from an IBM engineer.

Several lessons have been learnt from this evaluation. Firstly the results for Byblos are comparable with NSA's SIGINT experience (though admittedly somewhat worse), confirming that NSA's experience is applicable to our data.

Secondly this is the first time to our knowledge that the SAIL system has been objectively evaluated.

Thirdly it is the first time Attila has been trained on intercept. However there is a lot of uncertainty over the reasons for its worse performance than Byblos's. One factor,

probably, is lack of skill in its use: the IBM engineer who assisted Security Service was new to the field. Another factor is that experience from SIGINT applications has not fed into Attila in the way it has into Byblos. This was the interpretation BBN put on the result when informed of it: their lead developer commented that

> I doubt that IBM's fundamental technology is somehow irretrievably behind BBN's, but it's nice to know that the effort that you and we invest in making Byblos run "somewhat smoothly" on challenging data can pay off in this way.

Since this evaluation was completed, the IBM system has been retuned by IBM and the BBN system retuned by GCHQ (no further work has been done on the SAIL system). The current best performance is word error rate: BBN 60%, IBM 76%, SAIL 101% and word accuracy: BBN 45%, IBM 42%, SAIL 20%.

- **Bulk transcription**

The need for additional bulk transcription can be seen from the data presented in the Figure at the end of this report. It shows data points derived from NSA experiments on a variety of languages, as well as data points drawn from NIST evaluations sponsored by DARPA. Each point shows the measured word error rate (or character error rate for Korean and Mandarin) for a given number of hours of transcribed training data. All points are got using the Byblos system, and all except those labelled "DARPA English" correspond to experiments conducted on transcribed SIGINT data.

There are three lines drawn on the figure. The bottom one labelled "DARPA English" shows the performance of models built on public data, assessed on such data. There is a clear trend of improved performance associated with the use of more training data, but note that the improvement is only logarithmic.

The top one, labelled "Unclass. system on IA English" shows the performance of these same models on an Information Assurance application, where the speech to be transcribed is US English. The trend is the same, but there is a significant performance gap - of the order of 20 percentage points.

The middle line, labelled "IA English" shows the improvement that can be got by training a bespoke model for the task. There is still a substantial residual gap of around 7 percentage points between the DARPA line and the IA English line. The reason for this gap is not known, but it is clear that there has been a substantial improvement of performance – of the order of 13 percentage points – by using bespoke training.

The remaining points for other languages have much more variation, but overall are compatible with the existence of a similar trend of better performance associated with using more data. We have no information for these other languages on how much worse the performance would have been if public data had instead been used in the system training, these points are all drawn from models built using intercept.

The point for NIRAD English is high in comparison with the broad trend for all the non-IA English languages – one would have expected a word error rate of closer to 50% rather than the 62.5% measured. This may be due to the nature of the data, as it has been recorded with both sides of the conversation merged which is known to have an adverse effect on the performance of speech processing algorithms.

We cannot explain the substantial gap between the performance on IA English and that on all other languages; it may be attributable to an inbuilt bias in current speech recognition systems towards features of US English caused by decades of intensive research driven by US funding using US speech data.

GCHQ operational experience

GCHQ has been making operational use of Byblos since around 2004. The transcripts it produces unaided have not been of sufficient accuracy to have any value, but the technique of language-model biasing has enabled GCHQ to tailor Byblos to specific keywords or strings of interest. (The possibility of sharing techniques of this sort is a further reason to aim for compatibility between agencies.)

The first application was to strings of digits spoken by Caribbean drugs runners. GCHQ was able to detect spoken telephone numbers with high reliability using an out-of-the-box recogniser whose error rate was greater than 100% under the standard metric. Since then several instances of number detection have been deployed. In one recent case the digits are recognised with sufficient accuracy for it to be worth reporting their values to analysts, rather than just reporting their detection.

GCHQ has one deployed example of keyword detection other than spoken digits, but has had difficulty in persuading analysts to propose suitable search strings. GCHQ expects to be able to extend the range of deployments over the next couple of years, owing both to the wider range of languages available and to improved accuracy as Sigint corpora get transcribed. The operational benefit in the short term is likely to remain small compared with other technologies such as diarisation, gender and speaker ID.

Conclusion

The current state of technology is that systems are capable of automatic transcription with word error rates of between 30% and 40%, given amounts of training data of the order of hundreds of hours. The cost of transcribing this amount of training data is substantial – of the order of £0.5M for 300-400 hours of material.

The accuracy required of a system in order for it to provide business benefit will depend on the business application, and we do not yet have a good understanding of this. GCHQ have successfully deployed several STT applications to locate the

existence of spoken numbers such as telephone numbers in speech. They have also deployed a STT application which locates the existence of specific keywords.

In each of these applications, success has been achieved using an extremely poor core STT model (the default unclassified one supplied by BBN), with the performance enhanced by tailoring the language model. As the performance of STT systems improves, either by providing more training data or by technical advances in the algorithms used, so the range of applications for which they can provide business benefit will expand.

In the long term it is difficult to predict how the technology will evolve. Our judgement is that the recent improvement in performance driven by large-scale US investment is likely to plateau as the performance of STT on transcription of cooperative or public speech attains levels approaching 90% accuracy. US investment is now moving towards follow-on applications such as machine translation of the recognised speech.

There remains a significant gap between the performance measured on public data and the performance measured on intercept data, which may limit the potential for transcription of intercept data to accuracies of the order of 80% using current technology. However, to achieve such levels of accuracy will need substantial investment in bespoke training, and we should not wait for them to be achieved before seeking applications.

It is premature to choose between the IBM and BBN systems in terms of performance on classified material, as we only have one experiment to guide us. However the fact of the long experience of BBN in developing systems for use on SIGINT material makes it the preferred system for operational deployment in the short term.

State of the art speech recognisers are not shrink-wrapped products and require substantial training in order to understand how to use them and exploit them. There is no standard for STT models, and so models built for one recogniser are not portable to another. STT models are not cheap to build, requiring of the order of a year of CPU time (depending on the amount of data). These factors mean that there is considerable benefit to be had in UK agencies agreeing to use a common system in the long term, which would allow pooling of expertise and sharing of built models.

[redacted]

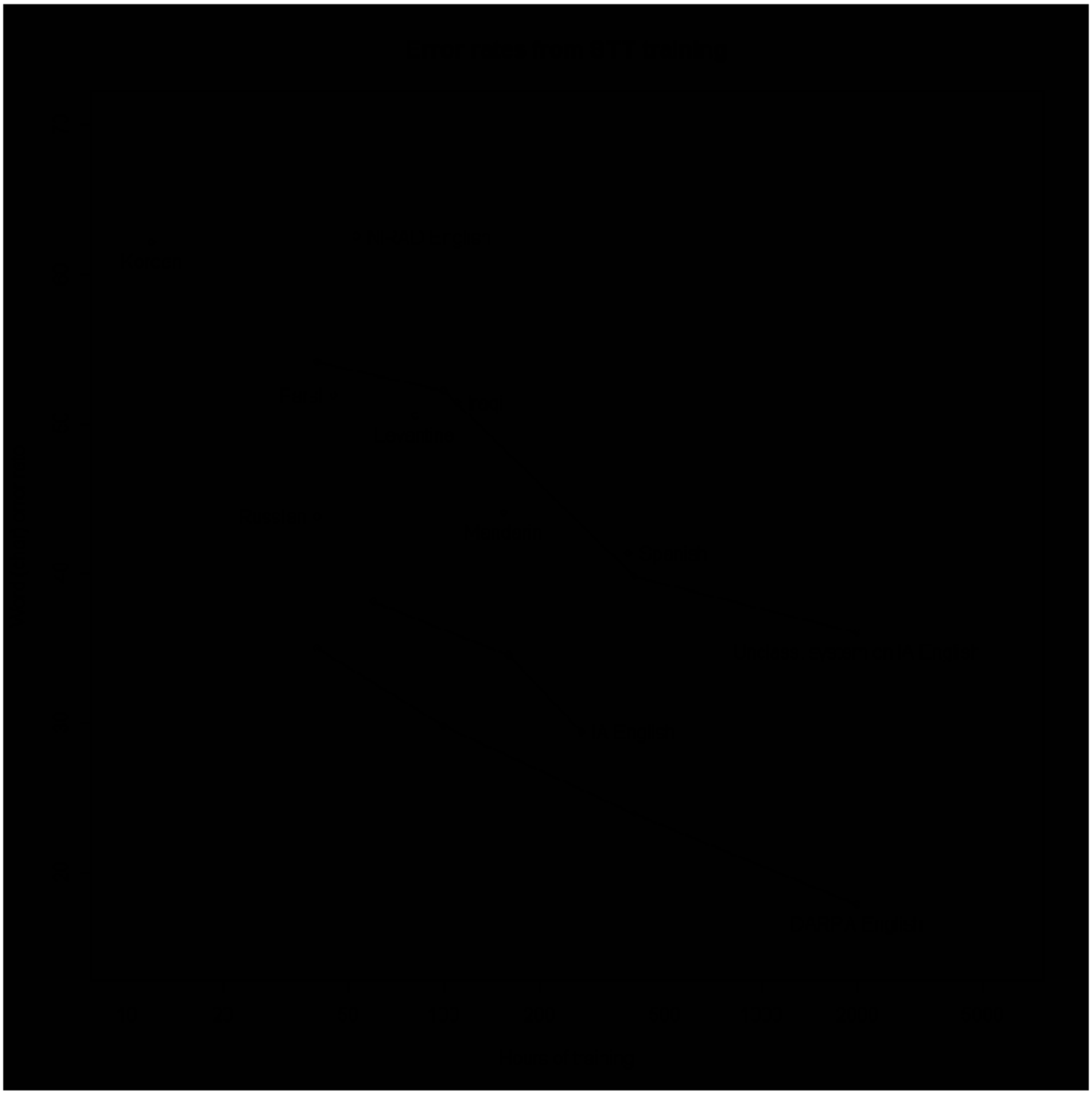
Chair, SIRDCC Speech Technology Working Group

References

[1] Minutes of SIRDCC Working Group Meeting on Speech Technology, 2001-12-03

[2] Comparative evaluation of three commercial speech recognisers, TRMCA/Inf/623, 2009-09-04 (revised 2009-12-07)

Figure: Error rates from training Byblos recogniser on different amounts of data



B/7655BA/1400/00006/018/0
7 December 2009

Appendix: Illustrative text and 100 well-recognised words

BBN Byblos transcription – correct words are marked in red

As delivered 2007

Truth: great o. k. that that's that's perfect o. k. well listen [talking] to derry give me i'll expect you there i will expect a call maybe some time thursday morning

Byblos: critical credit book books post post purple it was miles to go before you on the show communal experts will but the coma mission and mourn

Bespoke trained 2009

Truth: great o.k. *** that ** that's that's perfect o.k. well listen

Byblos: right o.k. but that is that's that's perfect o.k. **** what

Truth: [talking] to derry and [talking] give me i will expect you there i i will expect a call maybe some **** time thursday morning

Byblos: ********* on the fariones should give me * **** ******* ****** all to go to the hospital call maybe some cunt was a morning

The best-recognised words (other than 100%) with their frequency counts

94%		78%		73%		69%		66%	
CRAIC	17	SOMEBODY	18	LAST	26	NO	261	FELT	3
FUCKING	204	WEEK	22	PROBLEM	26	KNOW	390	FIFTY	15
SCALLY	9	FRIDAY	26	BELFAST	11	TOMORROW	45	FIND	12
MORNING	30	TWELVE	13	SIX	33	TOLD	35	HOPEFULLY	3
DIFFERENT	7	SEVEN	42	GIVE	76	NUMBER	57	JOB	9
MUMMY	14	AGAIN	29	RIGHT	284	[BREATH]	136	JOKING	3
NINETY	7	AIRPORT	8	TALKING	18	PHONE	47	LEAST	3
YEAH	339	ALREADY	4	REALLY	25	SAYS	135	MARATHON	3
WEEKEND	12	CHECKED	4	CHANCE	7	HALF	28	MOVING	3
BACK	103	DEAD	8	DRIVING	7	HUNDRED	86	MUCH	33
CLEAR	5	DUBLIN	4	ELEVEN	28	BEDROOM	3	NIGHTMARE	3
COUPLE	15	EACH	4	MOBILE	7	BLAME	3	OPPOSITE	3
DRINK	5	EXACTLY	8	PEOPLE	21	BRILLIANT	12	PASSPORT	3
KEPT	5	HOURS	8	NEXT	24	CHRISTMAS	6	PRESSURE	3
HELLO	100	KNOWS	4	BIG	17	CLEAN	6	PUB	3
COMING	19	LIVERPOOL	8	HOUSE	40	DATE	3	QUID	6
MINUTE	19	PARK	4	MONDAY	10	DERRY	3	SEAN	3

B/7655BA/1400/00006/018/0

7 December 2009

O'CLOCK	19	PICTURES	4	SOMEWHERE	10	DRINKING	3	SECONDS	3
DOUBLE	9	THIRTEEN	4	ANYWAY	23	DRUNK	3	SIXTY	9
REMEMBER	9	GRAND	15	TWENTY	36	DURING	6	SLOWLY	3

UK SECRET STRAP1