Vocal Analysis Software for Security Screening: Validity and Deception Detection Potential

Aaron C. Elkins, Judee Burgoon, and Jay Nunamaker

ABSTRACT

Our voices are encoded with emotional information. While it is complex and difficult to develop software to classify emotion and deception from the voice, it is possible. Using experimental methods, this research examines current commercial vocal analysis software for predictive and statistical validity in identifying emotion and deception for security screening. It is unrealistic to rely completely on the voice to detect deception and hostile intent for all people and all situations. But, by exploring the vocal variables used by the software, we are able to correspond and fuse them with other detection technologies for higher prediction reliability and accuracy. Implementing an unreliable and invalid detection technology could place the country's security in jeopardy by failing to detect actual threats. Just as deleterious, however, would be to dismiss technology, such as vocal analysis, before it has been thoroughly examined. This would deprive us of a valuable tool for detecting threats and securing our homeland.

INTRODUCTION

Imagine a time when a close friend or parent spoke to you. In the case of your parent, you knew immediately if they were angry or happy with you from their voice alone. Your parent spoke louder, faster, and in a higher pitch than usual after discovering you broke her grandmother's vase. Contrast this with a close friend who recently had a death in his family. He sounds depressed and speaks much slower and in a lower volume then an angry parent. With the thoughts of their loved ones on their mind people would sound distracted, with shorter responses and vocal interruptions. As social creatures, we can quickly and automatically determine emotional state or mood from the voice.

Despite how effortlessly we can interpret emotion and mood from the voice, developing computer software to replicate this feat is exceedingly difficult. Computers require very specific and predictable inputs and cannot deal well with unbounded contexts and the chaotic nature of conversation. We take for granted how complex conversations are and how quickly they branch and weave back and forth between topics and ideas. We even alternate between moods and emotions in just one conversation, from anger when recounting a mean boss to happiness when discussing an upcoming birthday party.

In addition to the complexity of conversation contexts, the science of measuring and classifying emotion and deception using the voice is in its infancy. Fear, for instance, is characterized as fast speech rate, higher mean pitch, low pitch variability, and lower voice quality.[1] However, the relationship between vocal measures and emotion has not been well explored beyond correlational analyses, leading to conflicting results and alternative vocal profiles for fear.[2]

Previous research has found that an increase in the fundamental frequency or pitch is related to stress or arousal.[3] Pitch is a function of the speed of vibration of the vocal chords during speech production.[4] Females have smaller vocal chords than men, requiring their vocal chords to vibrate faster and leading to their higher perceived pitch. When we are aroused our muscles tense and tighten. When the vocal muscles become tenser they vibrate at a higher frequency, leading to a higher pitch. Similarly, previous research has found that when aroused or excited, our pitch also exhibits more variation and higher intensities.[5]

Deceptive speech is also predicted to be more cognitively taxing, leading to non-strategic or leakage cues.[6] These cues, specific to cognitive effort, can be measured vocally. Cognitively-taxed speakers take longer to

respond (response latency) and incorporate more nonfluencies (e.g., "um" "uh", speech errors).

DISCUSSION

STATE THE PROBLEM

Despite the complexity of communication and the dearth of research in classifying emotion and deception from the voice, commercial software for automatically detecting emotion, stress, and deception is being adopted for use in law enforcement, fraud detection, and rapid screening environments.[7] Vocal analysis software is appealing because it provides a noncontact and inexpensive tool for rapid screening, requiring only a computer and microphone. However, most of the research on vocal analysis software focused on the older Vocal Stress Analysis (VSA) technology and not the current full vocal spectrum systems.

Investigations on modern full spectrum vocal analysis software found it unable to detect deception above chance levels.[8] However, all of this research examined the lie or truth classifications provided by the software interface and did not "look under the hood" at the underlying vocal measurements provided by the system and examine their validity and classification potential.

STATE THE POTENTIAL SOLUTION AND RESEARCH METHODOLOGY

This research investigates the validity and deception and emotion detection ability of commercial vocal analysis software using experimental methods. A series of experiments were conducted requiring participants to lie, commit a mock crime, and experience cognitive dissonance and stress. Participant's voices from each experiment were recorded and submitted to modern vocal analysis software for processing[9]. In addition to the classification provided by the software, the raw vocal variables were extracted from the software and analyzed using statistical and machine learning methods.

Replicating earlier research the vocal analysis software's built-in deception classifier performed at the chance level. However, when the vocal variables were analyzed independent of the software's interface, the variables documented to measure Stress, Cognitive Effort, and Fear significantly differentiated between truth, deception, stressful, and cognitive dissonance induced speech.

The results of a factor analysis suggest the existence of stable latent variables measuring Conflicting Thoughts, Thinking, Emotional Cognitive Effort, and Emotional Fear. A logistic regression model using the vocal measurements for predicting deception outperformed machine learning classification approaches (Support Vector Machine and Decision Tree) with a prediction accuracy ranging from 46 percent to 62 percent.

Despite the discouraging performance of commercial vocal analysis software's built-in classification, the variables underlying these classifications hold promise for predicting emotion and deception if properly calibrated to specific screening or security environments.

STATE THE END USERS/CUSTOMERS/ WHO WOULD BENEFIT

Since 9/11, the US Department of Homeland Security (DHS) has been seeking to increase the country's technological capability to secure its borders and airports. In response to this need a growing community of commercial security technology companies have emerged to service this niche industry. According to CBP officials, many of these vendors are "selling solutions in search of a problem." They may offer "one-size-fits-all technologies" with exciting feature lists. However, these systems depend on specific operating characteristics (e.g., polygraph style, rapid screening) and rely on single modalities (e.g., the voice).

This research investigates the potential of vocal analysis software to assist DHS in securing our borders and airports from threats.

STATE THE CHALLENGES TO ATTAINING THE SOLUTION AND RESULTS

The vocal analysis software vendors refute contradictory findings by arguing the built-in algorithms only work in the real world where tension, stress, and consequences are high. Creating these possibly harmful situations for experimental participants is not feasible. To overcome this limitation, careful experimental design based on communication and social psychology theory must be implemented to evoke emotions that occur during high stakes lies, without creating actual peril or harm.

If strong statistical relationships between vocal analysis software variables and emotions are replicated, we must try and interpret a black box system. The variables are calculated using propriety algorithms and are not standard. Research must occur in tandem corresponding these findings with standard phonetic measurements (e.g., f0, intensity, pitch contours) to better understand the emotional vocal behavior. This will further our scientific understanding and allow us to better calibrate vocal technology for specific security screening contexts.

We must be careful not to over rely on any one cue, vocal or otherwise. People and deceitful or truthful do not all behave the same. Some people may leak cues in their voice while others do not. Any technology solution implemented to observe and detect people should include multiple sensors. For the person that controls their voice well, their pupils, heart rate, or linguistic content will betray their hostile intent.

CONCLUSION

This research examines how reliable and valid commercial vocal analysis software is for predicting emotion and deception in security screening contexts using experimental methods. While research exists that evaluates current vocal analysis software's built-in classifications, there is gap in our understanding on how it may actually perform in a real high stakes environment.

Our voices are encoded with emotional information. While it is complex and difficult to develop software to classify emotion from the voice, it is possible. This research examines the variables produced by commercial vocal analysis software for predictive potential and statistical validity in identifying emotion and deception. It is unrealistic to rely completely on the voice to detect deception and hostile intent for all people and all situations. But, by exploring the vocal variables used by the software, we are able to correspond and fuse them with other detection technologies for higher prediction reliability and accuracy.

Implementing an unreliable and invalid detection technology could place the country's security in jeopardy by failing to detect actual threats. Just as deleterious, however, would be to dismiss technology, such as vocal analysis, before it has been thoroughly examined. This would deprive DHS of a valuable tool for detecting threats and securing our homeland.

ABOUT THE LEAD AUTHOR

Aaron C. Elkins is a postdoctoral researcher at the National Center for Border Security and Immigration (BORDERS), a DHS Center of Excellence at the University of Arizona. Aaron investigates how the voice and language reveal emotion, deception, and cognition for advanced human-computer interaction (HCI) and artificial intelligence applications. One application Aaron actively researches and develops technology for is automated interviewing and credibility assessment systems for rapid screening environments. These systems incorporate multiple behavioral and physiological sensors that inform an intelligent embodied conversational agent (AVATAR) interviewer. Complementary to the development of advanced artificial intelligence systems for security screening, is their impact on the people using them to make decisions. Aaron also investigates how human screeners are psychologically affected by, use, perceive, and incorporate the next generation of screening technologies into their decision making. He may be contacted at aelkins@cmi.arizona.edu.

¹ P.N. Juslin and P. Laukka, "Communication of Emotions in Vocal Expression and Music Performance: Different Channels, Same Code?" *Psychological Bulletin* 129, no. 5 (2003): 770-814; P.N. Juslin and K.R. Scherer, K. R., "Vocal Expression of Affect," *The New Handbook of Methods in Nonverbal Behavior Research* (2005):65–135.

² J.A. Bachorowski and M.J. Owren, "Vocal Expression of Emotion: Acoustic Properties of Speech are Associated with Emotional Intensity and Context," *Psychological Science* (1995): 219–224; Juslin and Scherer, "Vocal Expression of Affect."

³ Bachorowski andd Owren, "Vocal Expression of Emotion"; L.A. Streeter and others, "Pitch Changes During Attempted Deception," *Journal of Personality and Social Psychology* 35, no. 5 (1977): 345–350.

⁴ I.R. Titze and D.W. Martin, "Principles of Voice Production," *Acoustical Society of America Journal* 104 (1998): 1148.

⁵ Juslin and Laukka, "Communication of Emotions in Vocal Expression and Music Performance."

⁶ D.B. Buller and J.K. Burgoon, "Interpersonal Deception Theory," *Communication Theory* 6, no. 3 (1996): 203-242; P. Rockwell, D.B. Buller, and J.K. Burgoon, "Measurement of Deceptive Voices: Comparing Acoustic and Perceptual Data," *Applied Psycholinguistics* 18, no. 4 (1997): 471-484; M. Zuckerman, B.M. DePaulo, and R. Rosenthal, "Verbal and Nonverbal Communication of Deception," *Advances in Experimental Social Psychology* 14, no. 1 (1981): 59.

⁷ R. Holguin, "L.A. Co. Gets Cutting Edge Lie Detector," December 12, 2008, http://abclocal.go.com/kabc/story?section=news/bizarre&id=6554064.

⁸ K. Damphousse, L. Pointon, D. Upchurch, and R. Moore, *Assessing the Validity of Voice Stress Analysis Tools in a Jail Setting* (research report submitted to the U.S. Department of Justice, March 31, 2007); M. Gamer, H.G.Rill, G. Vossel, and H.W. Gödert, "Psychophysiological and Vocal Measures in the Detection of Guilty Knowledge," *International Journal of Psychophysiology* 60, no. 1 (2006): 76–87; D. Haddad, S. Walter, R. Ratley, and M. Smith, *Investigation and Evaluation of Voice Stress Analysis Technology* (Rome, New York: Air Force Research Lab Information Directorate, 2001), www.ncjrs.gov/pdffiles1/nij/193832.pdf.

⁹ The Nemesysco Layered Voice Analysis (LVA) 6.50 software was used to process all experimental recordings and produce the classifications and vocal variables used in the analysis.