The Voice and Eye Gaze Behavior of an Imposter: Automated Interviewing and Detection for Rapid Screening at the Border

Aaron C. Elkins
University of Arizona
aelkins@cmi.arizona.edu

Douglas C. Derrick
University of Nebraska at Omaha
dcderrick@mail.unomaha.edu

Monica Gariup
Frontex, Research and Development Unit
monica.gariup@frontex.europa.eu

Abstract

Contextual differences present significant challenges when developing computational methods for detecting deception. We conducted a field experiment with border guards from the European Union in order to demonstrate that deception detection can be done robustly using context specific computational models. In the study, some of the participants were given a "fraudulent" document with incorrect data and asked to pass through a checkpoint. An automated system used an embodied conversational agent (ECA) to conduct interviews. Based on the participants' vocalic and ocular behavior our specific model classified 100% of the imposters while limiting false positive errors. The overall accuracy was 94.47%.

1 Introduction

Unlike Pinocchio, liars do not exhibit universal behavior or physiological signals in all situations. Deception is often inappropriately reduced to either simply telling the truth or lying. However, there are many strategies for lying (e.g., omission, imposters, equivocation, hedging); situations where lying occurs (e.g., rapid screening, imposter, interrogation, conversation); varying consequences and power dynamics (e.g., parents, friends, boss, border guard, law enforcement); and interviewing styles (e.g., behavioral analysis interviewer, informal chat, guilty knowledge test, short answer format). All of these factors contribute to the type of behaviors and physiological responses that are exhibited and are theoretically expected. These contextual differences present significant challenges when trying to develop computational methods for detecting deception. In order to develop systems that can be used for reliable deception detection, we must constrain the complex problem of deception and manage the factors described above.

We conducted a field experiment with border guards from the European Union in order to demonstrate that by controlling some of the above factors and by developing context specific computational models, deception detection can be achieved robustly. In the experiment, some of the participants were given a "fraudulent" document with incorrect data and asked to pass through a checkpoint. An automated system used an embodied conversational agent (ECA) to conduct interviews. The system was equipped with vocal and ocular sensors, as well as an electronic passport reader. Based on the participants' vocal and eye gaze behavior a computational classification model was developed to identify imposters while limiting the number of false positives.

2 Embodied Conversational Agent

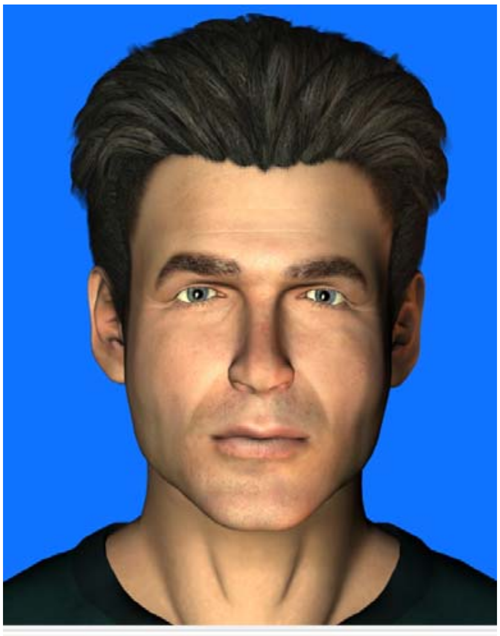
To account for the complex interplay between liars and the deceived, Buller and Burgoon (1996) introduced Interpersonal Deception Theory (IDT). This theory expanded and conceptualized deception as a strategic interaction between a sender and receiver. Liars must simultaneously manage information, their behavior, and appearance during the interaction. Moreover, liars will use different strategies depending on their skill, relationship with the interaction partner, preparation, motivation, and time.

Proceedings of the EACL 2012 Workshop on Computational Approaches to Deception Detection, pages 49–54,
Avignon, France, April 23 - 27 2012.

Lying is undeniably a social act. One major challenge to computational deception detection is accounting for the variability introduced by human interviewers. Every interviewer has their own style (e.g., aggressive, friendly), inconsistently asks questions, and gets tired. The behavior and approach of the interviewer strongly influences the behavior and reactions of the interviewee. For example, if the interviewer is angry, the interviewee will be affected by this and artificially display reciprocal anger or even distress. Perhaps after a lunch break the interviewer is fresh and in better spirits and returns to a more friendly interaction. Any deception detection system that relies on consistent behavioral cues will have to account for the diverse range of human interviewer variability.

To address this challenge, we developed an ECA-based deception detection system that asks the same questions, in the same order, and in the same way each time. Additionally, this system can speak the native language of every interviewee.

Figure 1. ECA Interviewer



3 Sensors

The ECA depicted above (Figure 1) conducts the structured border-screening interview and integrated into this system were three sensors for detecting imposters: microphone (vocalic measures), near infrared camera (ocular behavior), and an electronic passport reader (document input).

3.1 Vocalic Measures

A unidirectional microphone was integrated into the system to capture spoken responses to the ECA's questions. Vocal features were extracted from each of these responses near real-time (i.e., seconds). Previous research has found that an increase in the fundamental frequency or pitch is related to stress or arousal (Bachorowski & Owren, 1995; Elkins & Stone, 2011; Streeter, Krauss, Geller, Olson, & Apple, 1977). Pitch is a function of the speed of vibration of the vocal folds during speech production (Titze & Martin, 1998). Females have smaller vocal folds than men, requiring their vocal chords to vibrate faster and leading to their higher 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 (Juslin & Laukka, 2003).

Deceptive speech is also predicted to be more cognitively taxing, leading to non-strategic or leakage cues (Buller & Burgoon, 1996; Rockwell, Buller, & Judee K. Burgoon, 1997; Zuckerman, DePaulo, & Rosenthal, 1981). These cues, specific to cognitive effort, can be measured vocally. Cognitively taxed speakers take longer to respond (response latency) and incorporate more disfluencies (e.g., "um" "uh", speech errors). Moreover, the harmonics-to-noise ratio serves as an indicator of voice quality (Boersma, 1993). Originally intended to measure speech pathology (Yumoto, Gould, & Baer, 1982), liars have been found to speak with a lower harmonic-to-noise ratio than truth-tellers (Nunamaker, Derrick, Elkins, Burgoon, & Patton, 2011). The quality of the voice is affected by increased cognitive effort and heightened stress/emotion.

3.2 Ocular Behavior

This system was designed to be used in a rapid screening environment and to assess eye behavior during an interview typical at a port of entry. All participants were shown an image of his or her issued visa during the interview and asked if the information was correct. All of the information was correct on the visa for all participants except the imposters where the date of birth was inaccurate. This test design is based on orienting theory and predicts that measurable physiology accompanies an orienting reflex to familiar stimulus. Pavlov originally studied the orienting reflex during his classical conditioning experiments. This reflex orients attention to novel and familiar stimuli and is considered adaptive to the environment. In order to capture the eye behavior responses, we used the EyeTech Digital Systems VT2 infrared eye tracker (see figure 2) mounted directly below a computer monitor.

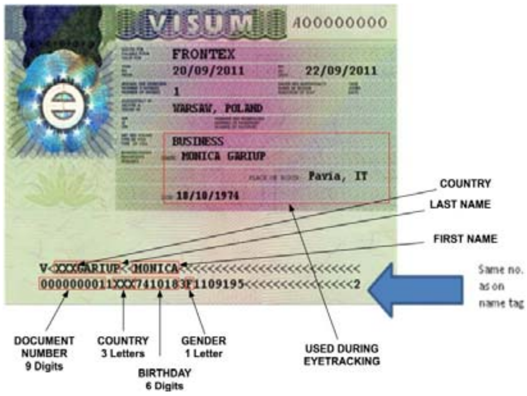
Figure 2. EyeTech Eye Tracker

The VT2 has two infrared light sources and an integrated infrared camera. It connects via USB to a Windows computer and captures the eye gaze location (x, y coordinates) at each instance at a rate of approximately 33-34 frames per second. During the interaction with the interviewing system, participants' eye behavior was monitored while they spoke to the ECA (e.g., for eye contact) and when they observed the image of their visa. Based on prior research (Derrick, Jenkins, & Nunamaker, 2011), we anticipated that the imposter would orient on areas of the image that contained false information about their identity. A sample of the document used by all participants is shown in Figure 3.

3.3 Electronic Passport Reader

To provide the system with additional information about each participant, a 3M AT-9000 e-passport reader was integrated into the system. Each participant placed their visa document on the scanner prior to the interview. The information from the document was read into the system using the Machine Readable Zone (MRZ) and an image of the visa was captured for use during interview.

Figure 3. Visa Issued to Participants



4 Final System

The final system used during the field experiment is depicted below in Figure 4. The ECA conducted the screening interview in the language of each participant's choice (English, French, German, Polish, or Spanish). The eye gaze and vocal behavior were recorded and processed during the interaction in near real-time. Immediately to the right of the monitor is a passport reader used to scan the entire participant's visas to read the information and image.

Figure 4. ECA System Screening Participants

5 Method

5.1 Participants

37 border guards from the following European Union (EU) member states participated: Austria, Belgium, Denmark, Estonia, Finland, Germany, Hungary, Italy, Latvia, Lithuania, Malta, the Netherlands, Norway, Poland, Romania, Slovakia, Spain, Sweden and the United Kingdom. In addition, two representatives from the U.S. Customs and Border Protection and the Canadian Border Services Agency also

participated. The simulation was part of a workshop on decision support for screening and risk assessment at passport control. Of the 37, four (two men from the Netherlands and Romania and two women from Sweden and Denmark respectively) were selected to be imposters with incorrect dates of birth on their visa. The selection was based on the results of an earlier survey that asked participants about their confidence in his or her ability to lie and remain undetected. Participants with the highest confidence scores were selected.

5.2 Procedure

All participants were given the visa document at their registration to use as their travel document during the simulated border check. The document consisted of personal data: name, surname, date of birth, place of birth, and nationality. For the sake of the simulation, all documents were assumed to be genuine (not counterfeited/falsified). The imposters were given basic instructions on the scenario and their role. They had acquired genuine documents of a person with a name identical to theirs who also looked like them (based on the photo of their own passport).

The goal of the imposter was to pass the border check undetected, meet with their fellow accomplices (the other imposters), and place a bomb at a reception. The imposters were explicitly instructed the day before the simulation to check his or her visa document and note that the date and location of their birth was false (not coincident with their real one).

The imposters were asked to lie and answer affirmatively when the ECA asked him or her whether the document belonged to them and if the data were correct. In order to reinforce the imposters in their role, on the day before the simulation, they were invited to meet in secret in the reception room and discuss how they would set up a bomb scare.

The other participants were simply told to behave normally. All other participants, and experimental staff were blind to the identity of the imposters.

Upon arrival to the in participants scanned their document in the e-passport reader and selected the language for the ECA to use for the interview (English, Polish, Spanish, French, or German). The ECA then asked the following questions:

1. What is your full name?
2. Where were you born?
3. Do have anything to declare?
4. What is the purpose of your visit?
5. Please carefully examine the following document
   a. Scanned document shown (15 seconds)
6. Are the data on this document correct?
   a. Scanned document shown (15 seconds)
7. Is this your document?

During each of the responses, the sensors monitored the vocal and ocular behavior of the participants and descriptive measures were calculated and presented to an experimenter and other participating border guard.

6 Results

6.1 Vocal Behavior

To assess the relationship between the voice and the participant's interview responses a multilevel model was specified with vocal quality as the response variable (N=189) regressed on condition (Guilty/Innocent) and question number (time). To reflect the repeated measure experimental design of multiple questions, both time and the Intercept of vocal quality were modeled to vary within Subject (N=38) as random effects. To calibrate each speaker each measurement of voice quality was subtracted by their starting value to reflect the deviance from a neutral starting point (Question One).

Table 1 below lists the fixed effects from this analysis. The imposters had a significantly larger drop in voice quality during the interview than innocent participants, $b=-2.18$, $p<.01$. Innocent participants and imposters both dropped their voice quality over time, likely because of the stress of the interview in contrast to the benign starting question.

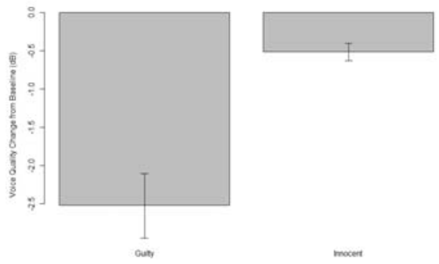
Table 1. Vocal Behavior Fixed Effects (N=189, 38 Subjects)

Fixed Effects	β
Intercept	-0.136
Imposter	-2.180**
Time (Question)	-0.134*

*p<.05; **p<.01; Fit by Maximum Likelihood Estimation.

The imposters spoke with much lower vocal quality when answering questions related to their identity (questions 5, 6, and 7 above). Figure 5 below illustrates the difference in vocal quality between innocent and guilty participants.

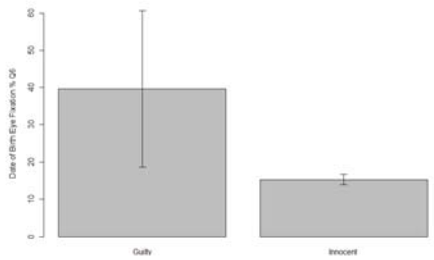
Figure 5. Lower Vocal Quality of Imposters



6.2 Eye Behavior

Eye contact with the ECA was not predictive of deception. However, a Between Subjects ANOVA revealed that when participants were asked to examine the image of their visa, imposters spent much more time fixating on the incorrect date of birth information, $F(1,36)=5.18$, $p=.028$, $\eta^2=.14$. Imposters spent 2.12 times more time fixating on the incorrect date of birth fields on their visa (Imposter M=29.73, SD=35.63; Innocent M=13.99, SD=8.45).

Figure 6. Eye Fixation Time on DOB Field



7 Summary of Predictors

To summarize the predictive potential of each of the collected vocal and eye gaze behavior measurements, each were submitted to Between Subjects ANOVA. Table 2 details the results for each measurement as a predictor of an imposter when answering question six.

Table 2. Vocal and Eye Gaze Behavior ANOVA Summary

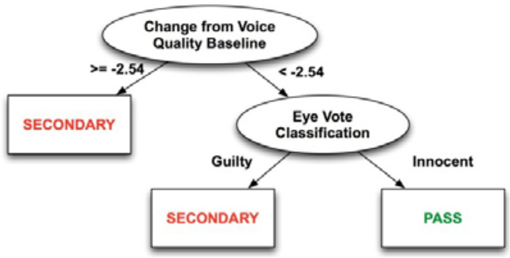
Predictor	df	F	p
Vocal Pitch Mean	36	0.05	.83
Vocal Pitch SD	36	0.30	.58
Vocal Quality Mean	36	8.78	<.01**
Vocal Quality SD	36	0.29	.59
Vocal Intensity Mean	36	1.65	.21
Vocal Intensity SD	36	0.82	.37
DOB Eye Fixation	36	5.18	.03*
Pupil Dilation	36	0.04	.83

*p<.05; **p<.01; DOB is Date of Birth field on visa document; All vocal measurements were speaker calibrated

8 Classifying Imposters

Vocal quality and date of birth fixation were submitted to a recursive partitioning classification algorithm (Clark & Pregibon, 1992; R Development Core Team, 2011). This type of classification algorithm has the advantage of being very easy to interpret and resulted in the decision rule detailed in Figure 7 below.

Figure 7. Imposter Classification Model



This final model had 94.47% accuracy, correctly identified all imposters, and misclassified two other participants of being imposters. When the classification model did not include the eye gaze behavior, the Voice Quality cut-off was much less conservative and resulted in many more false positives. However, after including the Eye Fixation variable, the system was calibrated to not over-rely on the voice.

This classification model illustrated the importance of additional sensors for improving overall accuracy of prediction, not just focusing entirely on true positives, or identifying imposters. Falsely accusing too many people would make the system infeasible in a high throughput, operational scenario. Given the diverse nature of the participants it should be noted that that gender, language, and potential cultural differences did not affect the results, but no support or conclusions can be drawn given the relatively small size of the various populations.

9 Conclusion

We conducted a field experiment with border guards from the European Union in order to demonstrate that by controlling some of the above factors and by developing context specific computational models, deception detection can be done robustly. We demonstrated that using both vocalic and ocular measurements we could correctly classify 100% of imposters in a limited scenario while limiting false positives. Future experimentation needs to be conducted to understand how the system compares to human judgment and if synergies exist between human and automated screening.

References

Bachorowski, J. A., & Owren, M. J. 1995. Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychological Science*, 219–224.

Boersma, P. 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences* (Vol. 17, pp. 97-110).

Buller, D. B., & Burgoon, J. K. 1996. Interpersonal deception theory. *Communication Theory*, *6*, 203-242.

Clark, L. A., & Pregibon, D. 1992. Tree-based models. *Statistical models in S*, 377–419.

Derrick, D. C., Jenkins, J., & Nunamaker, J. F. (2011). Design Principles for Special Purpose, Embodied, Conversational Intelligence with Environmental Sensors (SPECIES) Agents. *AIS Transactions on Human-Computer Interaction*, *3*(2), 62-81.

Elkins, A. C., & Stone, J. 2011. The Effect of Cognitive Dissonance on Argument Language and Vocalics. *Forty-Fourth Annual Hawaii International Conference on System Sciences*. Koloa, Kauai, Hawaii.

Juslin, P. N., & Laukka, P. 2003. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770-814.

Nunamaker, J. F., Derrick, D. C., Elkins, A. C., Burgoon, J. K., & Patton, M. W. 2011. Embodied Conversational Agent-Based Kiosk for Automated Interviewing. *Journal of Management Information Systems*, *28*(1), 17–48. doi:10.2753/MIS0742-1222280102

R Development Core Team. 2011. *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Retrieved from http://www.R-project.org/

Rockwell, P., Buller, D. B., & Judee K. Burgoon. 1997. Measurement of deceptive voices: Comparing acoustic and perceptual data. *Applied Psycholinguistics*, *18*(04), 471-484.

Streeter, L. A., Krauss, R. M., Geller, V., Olson, C., & Apple, W. 1977. Pitch changes during attempted deception. *Journal of Personality and Social Psychology*, *35*(5), 345–350.

Titze, I. R., & Martin, D. W. 1998. Principles of voice production. *Acoustical Society of America Journal*, *104*, 1148.

Yumoto, E., Gould, W. J., & Baer, T. 1982. Harmonics-to-noise ratio as an index of the degree of hoarseness. *Journal of the Acoustical Society of America*, *71*(6), 1544-1550.

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. 1981. Verbal and nonverbal communication of deception. *Advances in experimental social psychology*, *14*(1), 59.