



Deception Detection in Online Automated Job Interviews

Nathan W. Twyman¹(✉), Steven J. Pentland², and Lee Spitzley³

¹ Missouri University of Science and Technology, Rolla, MO, USA
nathantwyman@mst.edu

² Boise State University, Boise, ID, USA
spentland@cmi.arizona.edu

³ University at Albany, SUNY, Albany, NY, USA
lspitzley@cmi.arizona.edu

Abstract. This research-in-progress paper presents a conceptual system for automated deception detection in online interviewing. The design proposes video recordings of responses to predefined, structured interview question sets variously selected based on the desired behavioral metric of interest, such as competence, social skills, or in this case, veracity. Raw behavioral data extracted from video responses is refined to produce indicators of behavioral metrics. A prototype implementation of the design was built and tested experimentally using a job interview scenario. Results of the experimental analysis provide evidence of the potential of the concept.

Keywords: Behavioral assessment · Automated interviewing
Deception detection · Human risk assessment · Virtual agent-based interviewing

1 Introduction

Automated interviewing research is trying to bring information about potential human deception to more people in more places. Reliability has been a key hurdle to collecting and making use of deception assessments in job interviews, audit reviews, and other relevant applications. Identification of reliable human indicators of deception has been a major theme of this area of research. Linguistic, vocalic, oculometric, kinesic, and facial movement variations are among the sources of indicators that have been explored. Such human signal variations are referred to as *deception indicators* to the extent they are shown to be reliably correlated with deceptive but not truthful communication [1, 2].

Deception detection research has traditionally focused on identifying and exploring the robustness of specific indicators of deception, such as skin conductance or respiration rate. Automated human risk assessment research has followed a similar pattern, finding evidence for the potential of indicators such as movement freeze [3], pupil dilation [4], and vocal pitch [5].

Relatively little research has examined automated analysis of exploratory interviewing such as job interviews. The exploratory nature of these interviews precludes

exclusive use of short answer or binary yes-or-no questions, thereby falling outside the scope of some published design concepts [e.g. 6].

The knowledge proposition for this study is to outline concepts and constraints for a theoretical class of interview system that could derive behavioral metrics from open-ended interviewing scenarios, using an automated approach under the general framework of virtual agent-based interviewing [7, 8].

From the employer standpoint, the objective of the job interview is to determine if an interviewee is the most suited for a specific position. The interviewee in-turn seeks to present themselves in a way that makes the employer believe they are the best fit for the position. There is an expectation by the employer that only the most sought-after characteristics are being presented by the interviewee. And to some degree, there is an expectation that qualification enhancement or puffery is occurring. Levashina and Campion [9] found that 95% of undergraduate job seekers engage in some form of image creation or ingratiation when interviewing.

When an interviewee presents themselves to be more qualified than they are, they significantly increase the probability of making it to the next round of interviews [9]. Skillful self-presentation tactics may even be a quality that is necessary for specific types of work such as sales or customer service. However, interview performance does not necessarily correlate with actual job performance [10] – meaning that employers aren't getting the quality of worker that they expected based on the interview process. This can be especially problematic when companies seek specific technical skills and expect new-hires to possess a certain level of competency. When skill enhancement becomes exaggeration or fabrication, the long-term performance of a company can suffer when unqualified employees are hired.

2 Conceptual Development

The first property we propose for the concept is a structured questioning technique. Structured approaches to interviewing have been shown to curb the effects of self-presentation tactics [10], as interviewees simply have fewer opportunities to self-present in a manner that will sway the interviewer. Some IS research has used similar structured interviewing techniques to uncover concealed knowledge during screening interviews: Twyman et al.'s [6] *autonomous scientifically controlled screening systems* (ASCSS) class of systems use binary yes-or-no questions to remove variation due to question or response type from the interview process.

However, the interviewing scenarios that are the interest of this study require the use of open-ended questions. We propose that for these contexts, open-ended questions be generated in groups of three or more, with each group of questions possessing attributes similar in type, level, personalness, and importance. Each question set should contain questions of interest, but also at least two baseline questions, or in other words, questions for which responses are reasonably verifiable. Because they only request information that is already available or easily inferred, baseline questions are often excluded from these types of interviews. However, we propose they be purposely included for the current design.

For instance, in job interviews, a set of expected skills can be used as the basis for generating a set of interview questions about the interviewee's skills in that area. Some skills such as the use of basic office software might be so prevalent among applicants that they are virtually guaranteed. Yet questions about such skills could serve as the baseline questions in this design. If desired, greater certainty of the presence of these baseline skills could be obtained by pairing objective performance assessments with the automated interview.

Ultimately, open-ended responses to baseline questions should be compared to the questions of interest. For this stage, we propose taking an approach similar to prior IS research that has used non-invasive sensors to collect useful behavioral and psychophysiological data. Sensors used in this area of research include cameras (normal or specialized), microphones, platforms, and human-friendly lasers. In interview settings, these sensors have been used to generate raw oculometric, vocalic, linguistic, kinesic, proxemic, and even cardiorespiratory data, with varying degrees of fidelity [3, 4, 11–14].

This kind of data is subject to influence from many mechanisms besides deception, which is a key reason for using baseline questions. Large variations in behavioral responses to target questions as compared to baseline questions serve as flags or indicators of deception. Some related research has taken a data-driven approach when identifying indicators that are most diagnostic for the context of interest, while others have chosen a top-down theoretical approach. The former is excellent for discovery and exploration, while the latter provides greater confidence in reproducibility and generalizability.

In most cases, the theoretical explanation for automated veracity assessment has relied on the concept of leakage or strategic behaviors. *Leakage* theory asserts that lying produces natural human responses that the deceiver typically tries to mask or otherwise control. The theory suggests that such behavioral or psychophysiological responses “leak” out because of an inherent inability to control or mask them all [15]. Some later theories additionally proposed strategic behaviors as an explanation for some indicators. *Strategic* indicators are abnormal behaviors purposely exhibited by deceivers in an attempt to appear truthful [16]. Whether leaked or strategic, observed behavior when deceiving is compared to normal (i.e., truthful) behavior to gauge potential as a deception indicator [17].

The magnitude of these indicators varies from person to person. For instance, where one individual is naturally stoic in their speech or body movement, another may be naturally dynamic. A small drop in movement may be a major variation for one but minor for the other. Prior research has addressed this issue by requiring within-interviewee standardization prior to classification [6].

Neither leakage theory nor strategic explanations necessarily guarantee particular behaviors that will be displayed. Presumably, different deception indicators may be displayed depending on interpersonal, group, or cultural nuances. For instance, a person who is less worried about self-image may focus more on portraying believable linguistic content in their deception, while an image-conscious deceiver may put more effort into appropriate body language. With theories that provide no specific indicator guarantees, it is little wonder that no “Pinocchio's nose” or highly reliable indicator of deception has been found or is expected to be found. We therefore propose that an

effective design will necessarily measure and incorporate a breadth of potential deception indicators. A classification that fuses many indicators should more reliably catch deception when predicting which of many indicators will be present is not possible. Table 1 summarizes each property of the proposed class of systems.

Table 1. Summary of concepts for online interviewing system for deception detection

	Proposed system properties
1	Structured questioning technique, using homogenous question sets, each containing baseline questions and target questions of interest
2	Questions delivered by a virtual agent
3	During responses, collect raw psychophysiological and/or behavioral data using non-invasive sensor(s)
4	Scale data by interviewee prior to analysis
5	Fuse multiple disparate signals to generate a risk score reflecting deception likelihood

3 Method

To examine the potential of the proposed class of systems, we first instantiated the concepts in a prototypical system design. An experiment was conducted to evaluate the concept. Explanatory analyses were conducted to evaluate the potential of various behavioral indicators of deception. Performance analyses are currently in progress.

3.1 Prototype Implementation

To provide evidence toward proof-of-concept, we instantiated the design guidelines in an example prototypical system dubbed the Asynchronous Video-based Interviewing System, or AVIS.

AVIS displays interview questions sequentially in a text-based form, allowing respondents a limited amount of time to consider the question and a limited amount of time for a response. The amount of time for each is displayed to the interviewee between each question (see Fig. 1). Upon advancing, the interviewee considers the question for 60 s (see Fig. 2), then begins responding. Video recording does not occur except during response time (see Fig. 3), and this fact is made clear to the interviewee by showing a black screen where they would normally see themselves via a webcam.

Interviewee response recordings are tagged by question set, and stored for post-processing. AVIS extracts audio from each response, and extracts vocalic features from vocalic signal. Text is generated by applying IBM's Watson speech-to-text function to the audio signal, and linguistic features are generated by applying the text to SPLICE, a program that processes text and returns quantitative summaries of language and measurements of linguistic cues [18].

The video is currently processed using Intraface [19], a facial point mapping program, and Affectiva, a facial emotion classification system. These generate raw facial emotion measures and Cartesian coordinates of various points on the face for each frame of video.

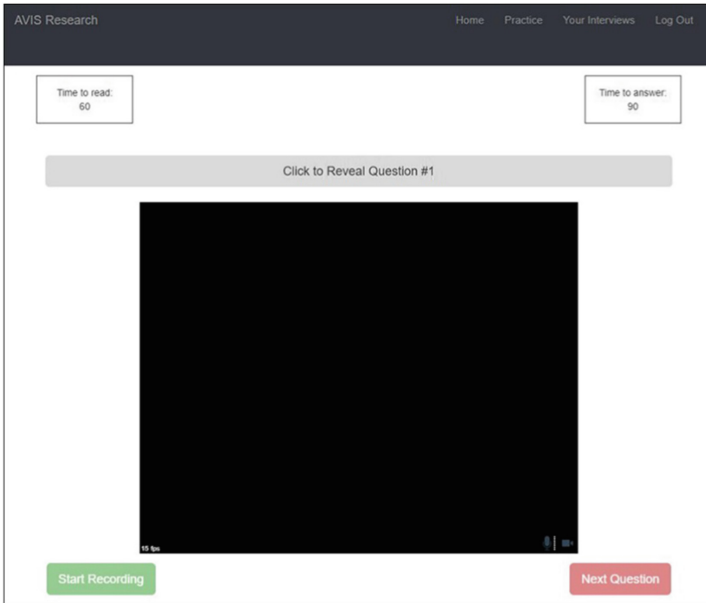


Fig. 1. Information screen displayed between questions

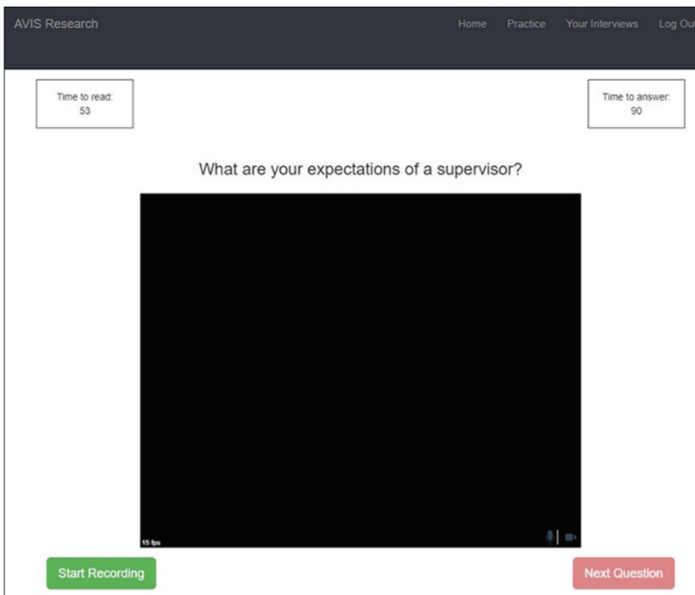


Fig. 2. Example question “pondering” stage, prior to response time. Interviewees have the option of starting the response whenever they are ready (via the green button)

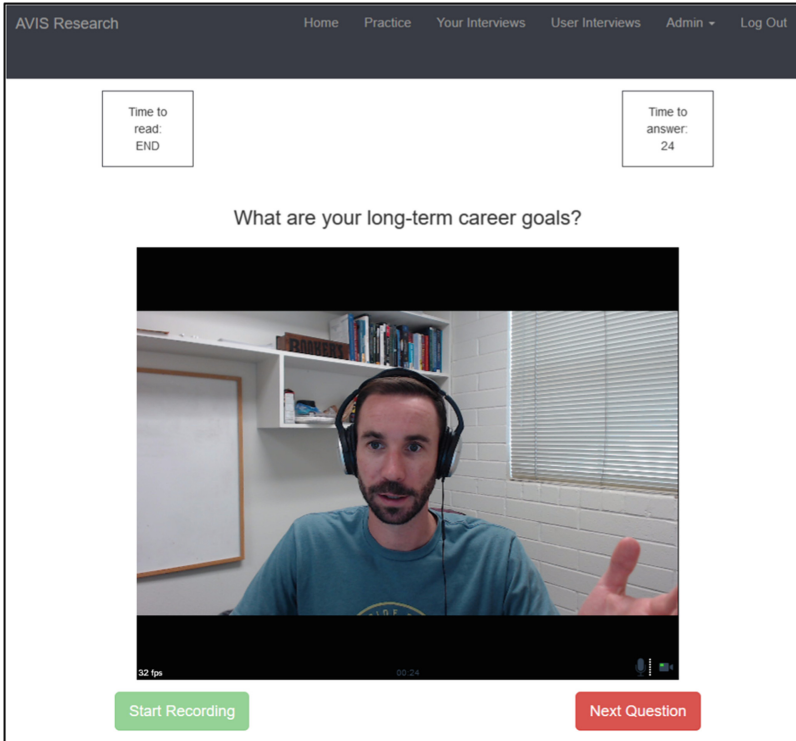


Fig. 3. Example interview question response

Whereas some prior research has used specialized hardware such as eye tracking systems and 3D cameras, such equipment is commonly not available in many potential application scenarios, and were not incorporated in this version of AVIS.

3.2 Experimental Evaluation

A mock job interview experiment was conducted with undergraduate students at a large university in the United States. The experiment employed AVIS as the interviewing mechanism for screening job applicants.

Experimental Procedure. Prior to arriving for the study, participants were asked to complete an online questionnaire. The questionnaire was designed to mirror a basic employment application and contained questions about education, work experience, and skills. The skill-related questions asked participants to rate their level of experience with common software packages, including Microsoft Excel, Microsoft Word, Adobe Photoshop, R Studio, and Oracle SQL. Additionally, participants were asked to rate their level of experience with StatView, a statistical package that does not exist.

When participants arrived for the study, they were told that they would be participating in a mock job interview using a one-way interview system. They were then

given a description of the job for which they would be interviewing. Participants were told that both their online application and interview responses would be evaluated to determine if they were suitable for the job. All participants were then allowed to “tailor” their original application to the job description. The participants were also told that if they were deemed to be a qualified candidate for the position, they would receive \$20. Otherwise, they would only receive \$5. This was a slight deception: all participants ultimately received \$20 for participation.

The job description listed several required skills, including “Advanced knowledge of Microsoft Excel” and “Proficiency with StatView Software Suite,” but did not mention many of the other skills outlined on the online application. The objective was to get participants to self-select to fabricate their qualification for specific skills to meet the requirements of the job description. Thus, there was no assignment to the Deception or Control condition; participants chose their condition.

While this self-selection is unusual for this type of research, we believe the self-selection bias in this case will be beneficial for evaluation of AVIS. Because of self-selection, the Deception group probably reflects those who are more comfortable with lying. The self-selection preserves a naturalness to the interaction that should elicit behaviors that reflect real-world lies.

Following the opportunity to tailor their online application, participants used AVIS to respond to 15 interview questions. The interview contained generic interview questions (i.e. “Tell me about yourself”) and a question set mirroring skills from the job description and online application. (i.e. “On a scale of 0 to 5 with 0 being none and 5 being a great deal, rate yourself level of experience with the following: Microsoft Excel Give a brief example to back your rating.”)

The design of the experiment allowed researchers to track exaggerations and fabrications made to an application to appear qualified for a job. The online application completed prior to viewing the job description was treated as ground truth. After the interview, participants took a post-task survey.

Participants. A total of 89 undergraduate students participated in the experiment (Male = 35; Female = 54). Of the total number of participants, 26% reported having previously used a similar one-way interviewing system during employment activities. When given the opportunity to tailor their online application, 43.8% of participants increased their self-rating for Microsoft Excel; 21.3% of participants increased their self-rating for Microsoft Word; and 31.4% reported having at least some level of experience with StatView, when previously reporting that they had never used the software.

3.3 Analysis and Results

Raw signals were averaged for each response. The averages for the responses to the skills questions were standardized within subjects to control for interpersonal differences. Video, audio, and linguistic data were separately submitted to principal components analysis for both key component identification and dimension reduction. A promax rotation was used because of the expected correlation between components. The final rotated components were labeled according to the behavioral trait they

seemed to be reflecting, based on the items that loaded heavily. The labeled components and example items are shown in Table 2.

Table 2. Labeled behavioral components

Vocal and linguistic components	Facial components
Vocal power trend	Mouth position
Vocal excitement	Facial affect
	Facial anger and contempt
Word complexity	Facial fear
Descriptive language use	Mouth openness
Language complexity	Main mouth movement
Present tense use	Upper eyelid movement
Word count	Bottom of nose movement
Negative word use	Nose bridge movement
Past tense use	Outer brow movement
	Inner brow movement
	Lower eyelid movement
	Mouth corners movement
	Main mouth acceleration
	Upper eyelids acceleration
	Bottom of nose acceleration
	Inner brow acceleration
	Outer brow acceleration
	Lower eyelid acceleration
	Mouth corners acceleration
	Nose to mouth corners vertical distance
	Inner brows to lower lids vertical distance
	Upper lips to lower lips vertical distance
	Upper lids to outer brows vertical distance
	Right outer brow vertical position
	Left outer brow vertical position
	Nostrils vertical position

A multivariate regression model was specified with Deception as the independent variable and the components in Table 2 as dependent variables. The linguistic variables were still undergoing data preparation and were not included. Deception had a significant impact on the twelve components displayed in Fig. 4. The units of measurement in Fig. 4 are standard deviations, so an interpretation for fear would be that on average, interviewees showed about half of a standard deviation more fear than normal when fabricating a response, as compared to their own baseline responses.

The results indicate that compared to the baseline questions, deceptive responses were associated with the total amount of movement dropping in locations across the face, and the acceleration of the facial movement that did occur was also slower. At the

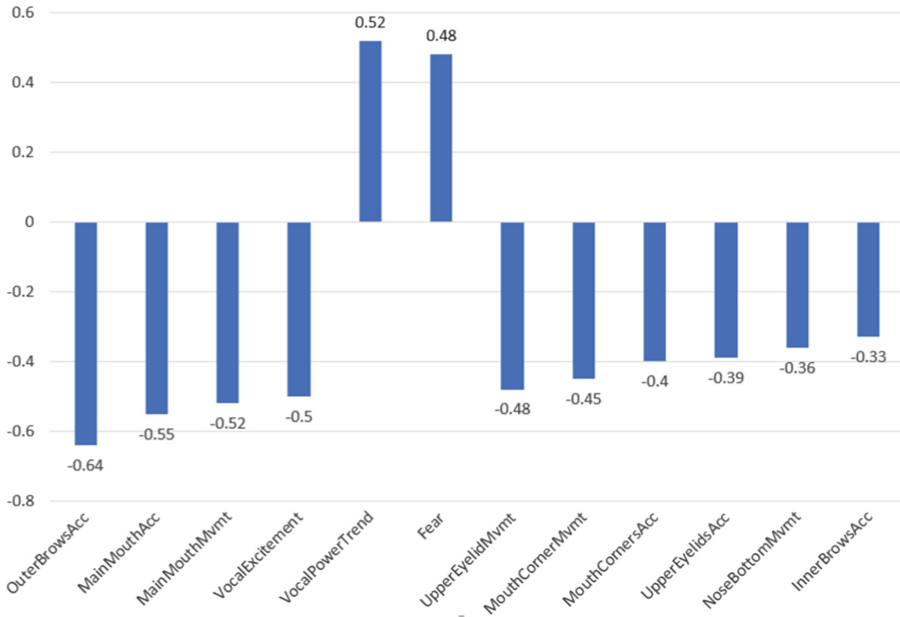


Fig. 4. Graphical representation of the coefficient estimates of significant ($p < .05$) behaviors during deceptive responses

same time, the amount of fear expressed on the face rose significantly. From a vocalic perspective, there was a significant drop in what we termed “Vocal Excitement,” which means the pitch, intensity and jitter of the voice decreased. Probably relatedly, the response marked a low point in a trend toward increased vocalic power in later responses.

4 Discussion

The goal of this study was to investigate the potential of a new class of system that could identify deception during interviews that required open-ended questions. Results of the experimental evaluation of the AVIS provide evidence that in this type of scenario, at least some behavioral variations do manifest that are diagnostic of deception. Specifically, facial animation decreases, both in terms of overall movement and the speed of change of that movement. Decreased facial movement has been identified as a potential deception indicator in a prior research study in a related context [20], and this study provides further evidence of this indicator, and indicates some level of robustness to a context with less control in terms of the allowance of open-ended questions. This is the first study identifying decreased facial acceleration as a potential deception indicator and its cause is unclear, but may be related to the facial freeze associated with both psychophysiological and behavioral mechanisms theorized to accompany deception. Additional research is needed to determine whether this is so.

Exaggeration. A follow-up investigation identified instances of exaggeration in the interview. While exaggeration is commonly identified as a type of deception, most of the behavioral indicators that were diagnostic of fabrication were not diagnostic of exaggeration. This is an important finding because some exaggeration is common in interviewing, and if such a system conflates it with fabrication, the value of its output may be diluted. Instead, the results suggest that exaggeration looks very different from fabrication when it comes to diagnostic behavioral indicators.

Additional analyses are underway. These focus on identifying the classification potential of this type of system but estimating AVIS's ability to predict deception. Though it is clear that deception creates behavioral anomalies, classification analyses will help provide an initial idea of how discriminatory and how useful the anomalies are.

References

1. DePaulo, B.M., Lindsay, J.J., Malone, B.E., Muhlenbruck, L., Charlton, K., Cooper, H.: Cues to deception. *Psychol. Bull.* **129**(1), 74–118 (2003)
2. Vrij, A.: *Detecting Lies and Deceit: Pitfalls and Opportunities*, 2nd edn. Wiley, London (2008)
3. Twyman, N.W., Elkins, A.C., Burgoon, J.K., Nunamaker Jr., J.F.: A rigidity detection system for automated credibility assessment. *J. Manag. Inf. Syst.* **31**(1), 173–202 (2014)
4. Twyman, N.W., Proudfoot, J.G., Schuetzler, R.M., Elkins, A.C., Derrick, D.C.: Robustness of multiple indicators in automated screening systems for deception detection. *J. Manag. Inf. Syst.* **32**(4), 215–245 (2015)
5. Derrick, D.C., Elkins, A.C., Burgoon, J.K., Nunamaker, J.F., Zeng, D.D.: Border security credibility assessments via heterogeneous sensor fusion. *IEEE Intell. Syst.* **25**(3), 41–49 (2010)
6. Twyman, N.W., Lowry, P.B., Burgoon, J.K., Nunamaker Jr., J.F.: Automated screening for detecting purposely concealed knowledge in individuals. *J. Manag. Inf. Syst.* **31**(3), 106–137 (2014)
7. Derrick, D., Jenkins, J.L., Nunamaker Jr., J.F.: Design principles for special purpose, embodied, conversational intelligence with environmental sensors (SPECIES). *ACM Trans. Hum. Comput. Interact.* **3**(2), 62–81 (2011)
8. Nunamaker, J.F., Derrick, D.C., Elkins, A.C., Burgoon, J.K., Patton, M.W.: Embodied conversational agent-based kiosk for automated interviewing. *J. Manag. Inf. Syst.* **28**(1), 17–48 (2011)
9. Levashina, J., Campion, M.A.: Measuring faking in the employment interview: development and validation of an interview faking behavior scale. *J. Appl. Psychol.* **92**(6), 1638 (2007)
10. Barrick, M.R., Shaffer, J.A., DeGrassi, S.W.: What you see may not be what you get: relationships among self-presentation tactics and ratings of interview and job performance. *Am. Psychol. Assoc.* **94**(6), 1394–1411 (2009)
11. Burgoon, J., Nunamaker, J.F., Metaxas, D.: *Noninvasive measurement of multimodal indicators of deception and credibility*. University of Arizona (2010)
12. Burgoon, J.K., Jensen, M., Twyman, N.W., Meservy, T.O., Metaxas, D.N., Michael, N., Elder, K., Nunamaker Jr., J.F.: Automated kinesic analysis for deception detection. In: *Proceedings of the HICSS-43 Symposium on Credibility Assessment and Information Quality in Government and Business*, Koloa, Hawaii, USA, 5–8 Jan 2010

13. Meservy, T.O., Jensen, M.L., Kruse, J., Burgoon, J.K., Nunamaker, J.F.: Automatic extraction of deceptive behavioral cues from video. In: Kantor, P., Muresan, G., Roberts, F., Zeng, D.D., Wang, F.-Y., Chen, H., Merkle, R.C. (eds.) *ISI 2005. LNCS*, vol. 3495, pp. 198–208. Springer, Cham (2005). https://doi.org/10.1007/11427995_16
14. Moffitt, K., Elkins, A.C., Burgoon, J.K., Nunamaker Jr., J.F.: Rapid noncontact credibility assessment via linguistic/vocalic analyses. In: *Proceedings of the National Communication Association 96th Annual Convention*, San Francisco, CA, USA (2010)
15. Ekman, P., Friesen, W.V.: Nonverbal leakage and clues to deception. *Psychiatry* **32**(1), 88–106 (1969)
16. Buller, D.B., Burgoon, J.K.: Interpersonal deception theory. *Commun. Theory* **6**(3), 203–242 (1996)
17. Elkins, A.C., Derrick, D.C., Gariup, M.: The voice and eye gaze behavior of an imposter: automated interviewing and detection for rapid screening at the border. In: *Conference of the European Chapter of the Association for Computational Linguistics*, Avignon, France (2012)
18. Moffitt, K.C., Giboney, J.S.: Structured Programming for Linguistic Cue Extraction (SPLICE). <http://splice.cmi.arizona.edu/>
19. De la Torre, F., Chu, W.-S., Xiong, X., Vicente, F., Ding, X., Cohn, J.F.: Intraface. In: *IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, Ljubljana, Slovenia (2015)
20. Pentland, S.J., Twyman, N.W., Burgoon, J.K., Nunamaker Jr., J.F., Diller, C.B.R.: A video-based screening system for automated risk assessment using nuanced facial features. *J. Manag. Inf. Syst.* **34**(4), 970–993 (2017)