2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)

Temporal Patterns of Facial Expression in Deceptive and Honest Communication

Taylan Kartal Sen
University of Rochester
Rochester, NY, United States
tsen@cs.rochester.edu

***Abstract*—Video recordings 151 online dyadic conversations involving individuals playing a deception-based interrogation game were collected for analysis of temporal features involving facial expression. The interrogation game involved two participants (one *interrogator* and one *witness)* engaging in an interrogation in which the *witness* was directed to either lie or tell the truth regarding *evidence* that was presented to the witness. Analysis identified that interrogators paired with dishonest witnesses demonstrated a higher variance in their smile-related lip corner puller action unit expression than interrogators paired with honest witnesses. We propose to apply HMM-based multimodal models to identify the temporal patters in facial expressions in dyadic communication and hope to discover differences in such patterns between honest and dishonest communication.**

1. Introduction

Ancient Indian papyrus texts (The Vedas), from as early as 900 BC, introduced the notion that a person's internal state of mind is unconsciously revealed by his/her external behavior. The Vedas give instruction to aid in the detection of poisoners: "[The poisoner] … does not answer questions, or they are evasive answers; he speaks nonsense … his face is discolored …". [1] Facial expressions were also analyzed by Charles Darwin, in his 1872 book, *The Expression of Emotions in Man and Animals*, stating that "...actions become habitual in association with certain states of the mind, and are performed whether or not of service in each particular case..." [2]. This premise was extended more recently by Paul Ekman and Wallace Friesen in which they promote the theory that nonverbal cues are more revealing than verbal behavior in determining deception [3]. However, despite extensive investigation by many researchers in both nonverbal and verbal behavior, today's specialists trained in behavior detection at the Transportation Security Agency (TSA) have been largely unsuccessful at detecting travelers associated with terrorism [4,5]. Indeed, numerous studies have shown that human accuracy in detecting deception is approximately 54%, even for trained experts [6,7,8]. Advancing our understanding of deception will not only aid national security, but also may be of use in reducing harm in several other ways, including fraud. The ability understand deception in an online communication medium is of particular interest with our the increasingly Internet-connected world.

In this paper, we investigate the temporal characteristics of facial expression in 151 pairs of videos of individuals playing an online game involving interrogation with both honest and deceptive communicators. We propose several strategies for identifying patterns in facial expressions and present some early findings. In particular, we try to address the following research questions:

1. Are the rates of change of particular facial features different between honest and dishonest communicators?
2. Is unsupervised clustering of facial expression measures useful in identifying patterns and reducing data dimensionality?
3. Is a Hidden Markov Model a good model for representing an individual's hidden mental state, and the observed facial expressions during dyadic conversation?

2. Background

Numerous studies have been conducted on lie detection involving a varied range of observational input features from thermal camera images to functional MRI scans, and diverse set of analysis techniques. In this section we place primary focus on nonverbal features which can be obtained from a simple web camera, nonverbal features, and studies focusing on dyadic data.

Yu, et al. looked specifically at the level of synchrony in head nodding between a witness and an expert interrogator as an indicator of deception [9]. Using N=121 video dyads, Yu et al. used automated head posture analysis to detect head pose and looked at the correlation between witness and interrogator head pitch. Yu's work is based on the broad communications theory of Interpersonal Deception Theory, which posits that the level of synchrony between a message sender and message receiver goes down with the message sender is being deceptive and found support for this theory in their study. Lu, et al. examined hand and head movements to understand deception [10], and Caso, et al., found that rhythmic gesturing is more characteristic of deceptive communicators [11]. Cohen, et. al. found that deceptive individuals used gestures which contradicted their speech [12].

Rosas et al. used a multimodal approach to deception detection using numerous set of features, including a number of facial features, including smile [13]. Facial expressions relating to happiness have also been shown to be associated with high stakes lies in televised statements by individuals pleading for the return of their missing family members [14].

There has also been research on the dynamic characteristics and temporal patterns of facial features in deception. The concept of micro expressions, patterns lasting less than 1/5 of the second, which reveal suppressed emotion, were investigated by Ekman et al. [15]. The timing of smiles specifically has been shown to be important to distinguish genuine delight from frustration by Hoque, et al. [16].

Dyadic facial features were investigated using Hidden Markov Models ("HMMs") by Li, et al. with regards to negotiation [18]. An HMM using facial features from both parties of a dyadic video chat based negotiation was used to model temporal properties and mutual influence between parties in order to predict the negotiation outcome.

3. Scope of the Technical Problem

Despite the attention deception detection receives from research and industry, expert systems have been unable to demonstrate repeatable classification accuracies at a high level [5]. Indeed, in United States courts of law, lie detection technologies, including polygraph, voice analysis, and MRI based systems, have consistently failed to meet a minimum level of validity to be admissible.

There are many reasons why detecting deception may be hard. Human thoughts and emotions are complex. Individuals behave differently, and two people thinking the same thoughts or being in the same mental-emotional state may have very different externally observable behaviors.

Additionally, the context in which an individual is communicating is likely to drastically affect his/her external behavior. Because of the large variation per individual and context, gathering a large amount of data involving deception is crucial. Laboratory studies are limited in scale in terms of location resources as well as staff needed to conduct studies. However, with the vastly increasing access of individuals with online video connectedness, and the growing poll of crowd-sourced workers, the idea of an automated online system for gathering research data is compelling. Our effort began with an attempt to develop a semi-automated system for recruiting and pairing participants, and directing them to take part in a communications study protocol. In developing the communication study protocol, special attention should be made to extract features that will be useful in accounting for the effects personal differences.

In addition to gathering data in a non-automated fashion, it is difficult to hand-code video and audio data. Hand coding data may also be more likely to introduce researcher bias. Thus, a second step in our work is to identify automated analysis methods for scalably conducting deception research. Only with substantially more data may we begin to apply more advanced models and machine learning methodology.

Our first step has been the creation of an online system helping to automate the gathering of dyadic data. The system was designed to gather data with limited researcher intervention on a dyad by dyad basis. Additionally, the system was designed to allow for ease of generating different testing contexts in order to easily obtain data in controlled scenarios.

4. Methods and Proposed Solution

In order to study nonverbal behavior associated with deception we first developed a system for automatically gathering dyadic data. While generic options for online video calling and recording existed [18], no system existed for simultaneously directing participants to "play through" our desired interrogation questioning protocol while recording video at a high quality. By building a framework instead of a single system, we provide others the ability to rapidly develop new systems with a modified study protocol and/or context. In order to take advantage of the scalability of online services and the real-life setting of participants in his/her own home environment, we developed our framework as an Internet-based platform. In order to overcome the problem of inconsistent quality online workers and questionable quality video, our framework incorporates quality control as a fundamental element of the final system.

4.1. Interrogation Questioning Protocol

Our dyadic video data was obtained with an interrogation protocol specially designed to allow us to address the person to person differences in expression. The data was gathered using an online web application system we developed which recruits crowd-sourced workers and directs them to play the game with another participant via video chat which is recorded. The system directs the participants through each step of the protocol outlined in Figure 1.

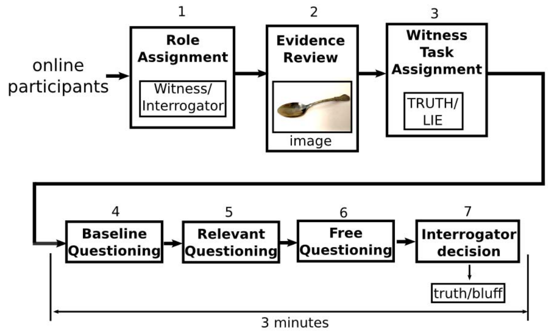

Figure 1. Steps of the Interrogation Protocol

The protocol begins by randomly assigning one participant the role of *witness* and the other participant the role of *interrogator* in the Role Assignment step. In the following Evidence Review step, the witness is shown a randomly selected picture from a set of detailed yet common item images (e.g. a spoon). The system directs the witness to try to remember as many details as they can about the image during a review period of 30 seconds. Next, in the Witness Task Assignment step, the online system randomly decides the witness's role of either telling the truth or bluffing. Specifically, for telling the truth the witness is told to describe the evidence image as truthfully as possible. For bluffing witnesses, the online system directs the witness to bluff and tell the interrogator his/her evidence image was X, where X is a one or two word name of a different object, randomly selected from the same group as the set of images. In the Baseline Questioning phase, the interrogator is prompted to ask the witness the baseline questions listed in Table 1 below one by one. Next, in the Relevant Questioning phase, the interrogator is prompted to ask the witness specific questions about the image. The Relevant questions sequentially vary from broad to specific in their level of detail asked. In the Free Questioning phase, interrogators are allowed to ask his/her own questions. The participants were each offered $10 to play the game and potentially received a $5 bonus. A $5 bonus was offered to the interrogator if they could correctly determine if the witness was lying or not. A $5 bonus was paid to the witness if the witness could get the interrogator to believe them (regardless of whether the witness was lying or telling the truth). The online system captured not only audio and visual recordings of the participants, but also recorded the interrogator's decision. Additionally, the system recorded the time for each time the interrogator pressed a "next question" button on the system's user interface.

The baseline questions in this protocol were designed to help account for individual variability in the type of facial expressions that they show in response to common mental states. Since, even though some aspects of facial expression may be to some extent universal (e.g. smile generally associated with friendliness in many cultures), expressions are likely to vary significantly from individual to individual. More specifically, our protocol includes a baseline questioning phase, based on the common polygraph Control Question Test, in which the witness is asked questions unrelated to the true matter the interrogator wishes to determine. This allows us to "normalize" for a person's individual variability in facial expressions from universal norms. Shown in Table 1 are the baseline questions our online data gathering system directs interrogators to ask his/her corresponding witness. The questions were picked to specifically probe an individual's facial expressions associated with the mental states of confusion, memory recall, analytical thought, discomfort, or simple question answering.

Table 1. Baseline Questions & Evoked Mental State

Baseline Question	Evoked Mental State
Am I wearing glasses or not?	simple question normal mental state
What type of computer system are	ambiguous question slightly confused state
What color clothes did you wear	memory recall
What is 12 + 19?	analytic thought
Did you ever steal anything in your whole life and if so what was it?	discomfort

By including these baseline questions, we thus are able to create a personalized template of an individual's expressions for a number of common mental states to be used in later analysis.

4.2. Analysis Methods

We used both off-the-shelf feature extraction tools as well as our own developed algorithms for extracting low level features from the raw audio-video files. For automated facial expression extraction we used OpenFace, an open source tool which has been benchmarked against several public databases [19]. Specifically OpenFace extracts facial action units, a well established facial movement coding methodology in which movements are tied to a single or small number of facial muscles. The OpenFace tool specifically has two types of output for each facial action unit, a Boolean output indicating expression or non-expression of the particular action unit, as well as a numerical output providing a measure of the intensity of expression. OpenFace also extracts the participant's head pose from the videos. The dynamic activity level of each extracted facial action units is calculated by taking the absolute value of the first derivative of the numerical OpenFace outputs.

An additional temporal feature which pilot testing suggested was important is the delay between the moment an interrogator completes a question and the moment a witness starts to respond ("response delay time"). We surmised that features related to this measure would show differences between honest and dishonest communicators, notably, that dishonest individuals would take longer to respond.

Based on past groups' success in using verbal features to detect deception, we plan to use several word embedding methodologies. The VADER sentiment analysis tool is likely to be of particular importance in deception detection due to the close association for facial features, emotion, and lying [20]. Another tool, the Linguistic Inquiry Word Count ("LIWC") may indicate differences frequency of subject matter that differs between honest and dishonest communication [21].

Our proposed solution to identify differences in temporally features started with the development of a system which will aid in the gathering of larger data sets. The above listed features will represent the intermediate data which will base our higher level models on. We plan to apply HMM and other probabilistic models to these features in order to characterize honest and dishonest communication and highlight any differences between the two.

5. Work Completed and Future Plan

5.1. Initial findings from Developed System

We used our semi-automated online data collection system to gather 151 dyads of audio-video recordings of participants playing the interrogation game. As shown in table II, initial statistical analysis of the variance in smile related facial action units AU 12 (lip corner puller) did not show substantial differences between honest and dishonest witnesses. However, significant differences were identified in interrogators, with the interrogators who were paired with dishonest witnesses smiling more dynamically than those who were paired with honest witnesses.

Table 2 – Differences in the mean variance of lip corner puller facial unit in truthful and bluffing witnesses and interrogators

Feature	mean variance for Truthful	mean variance for Bluffing	t-test
interrogator AU 12	0.101	0.139	**0.005**
witness AU 12	0.210	0.132	0.129

Initial analysis of response delay times between honest and deceptive witnesses did not show any substantial differences.

5.2. Future Planned Analysis

In order to reduce the high dimensional space of possible facial action units (17 from OpenFace, which makes for 2^17 = ~128,000 faces) to a space that is more easily managed (given the 1.3 million frames of data), we propose to use unsupervised clustering. By clustering the observed faces, we identify which action unit combinations are more likely, and can identify how many clusters are needed to do a *good* job of representing the data.

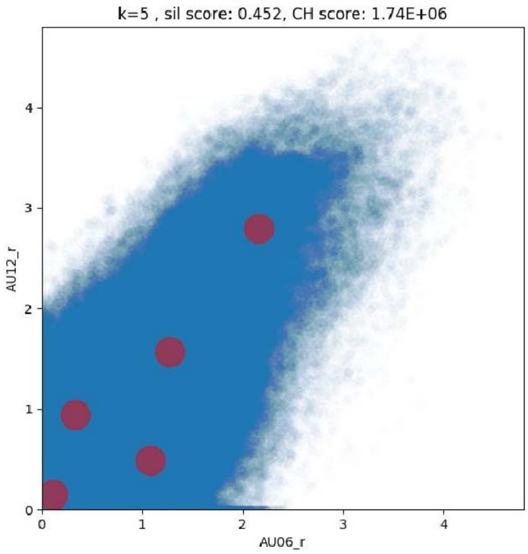

Figure 2. Clustering of AU 6 and AU12

Initial results of applying k-means clustering to the smile-related action units' (AU6 and AU12) continuous data outputted from OpenFace for all witnesses and interrogators is shown in Figure 2. The Silhouetting methodology was used to identify the ideal number of clusters, which was k=5. The resulting clusters represent interesting values for AU 6 and AU12. Shown in Figure 2, the bottom left cluster appears to correspond to a neutral face in which there is low cheek raiser and lip corner puller expression levels. Along the diagonal there appear two clusters which appear to both represent "true" or "Duchenne" smiles, in which both the cheek raiser and lip corner puller expressions are equally present. The cluster on the left above the origin appears to be a "fake" or "polite" smile in which only the lip corner puller is present. It is notable that unsupervised clustering was able to automatically discern that when looking at levels of smile related facial features people tend to express 5 types of smiles. The data reduction from a five dimensional continuous space to a 5 valued single variable represents a substantial reduction in complexity.

We plan to use these clusters as the observed outputs of an HMM. The hidden states, we conjecture, will apply to mental states. Our plan is that the transition probabilities in the mental states will provide the computational machinery necessary to identify and capture patterns that are important in distinguishing deceptive and honest communication.

We additionally plan to make use of the baseline questions corresponding to the various mental states in training the HMM. Thus, instead of completely hidden states for our model, it will contain both hidden states, as well as the set of mental states probed in the baseline questions including memory, analytical, confusion, and discomfort.

6. Research Contributions

This future work will significantly advance the scientific foundations and methodologies for temporal modeling of multimodal facial and verbal features. As discussed in section 5, our analyses have already shown that an interrogators variance in lip corner puller expression is a useful indicator for discerning honest and dishonest communication. We are confident that by utilizing an individual's expression during specific baseline mental states (corresponding to the analytical, memory, and discomfort-based baseline questions) improved analysis will result.

We believe we are the first to propose the use of HMM models based upon personalized facial expressions obtained during specialized baseline questioning. We hope the such models combined with the use of our semi-automated dyadic data capture system will provide significant advancement to the field of studying deception.

Conclusion

In this paper, we demonstrate that the our online system for semi-automated dyadic audio-video capture is an effective system for studying deception by gathering 151 high quality dyads which demonstrate significant findings. Initial findings include identification that interrogators speaking with dishonest witnesses tend to have more dynamic smile-related action unit expression. We hope that the proposed future work involving use of individual expressions during the various mental states probed during baseline question, clustering, and HMMs, will help contribute to new findings in the characteristics of honest and deceptive communication.

References

[1] Trovillo, "History of Lie Detection", Amer. J. of Police Science, 1939, p849.

S. Weinberger, “Intent to deceive ?,” Nature, vol. 465, no. May, pp. 412–415, 2010.

[2] Charles Darwin. *The Expression of the Emotions in Man and Animals*, D. Appleton And Company, 1899, (first edition was published in 1872).

[3] Paul Ekman and Wallace V. Friesen. “The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding” Semiotica, 1969, I(I), 49-98.

[4] R. GAO Accountability, Integrity, “Efforts to Validate TSA ’ s Passenger Screening Behavior Detection Program Underway , but Opportunities Exist to Strengthen Validation and Address Operational Challenges,” Gao, no. May, 2010.

[5] Weinberger, “Intent to deceive ?,” Nature, vol. 465, no. May, pp. 412–415, 2010.

[6] M. G. Aamodt and H. Custer, “Who can best catch a liar?,” The Forensic Examiner, vol. 15. pp. 6–11, 2006.

[7] C. F. Bond and B. M. Depaulo, “Accuracy of Deception Judgments,” vol. 10, no. 3, pp. 214–234, 2006.

[8] DePaulo, B.M., Lindsay, J. J., Malone, B. E., Muhlenbrick, L., Charlton, K., & Cooper, H. Cues to deception. Psychological Bulletin, 2003

[9] Yu, X., Zhang, S., Yan, Z., Yang, F., Huang, J., Dunbar, N.E., Jensen, M.L., Burgoon, J.K. and Metaxas, D.N., 2015. Is interactional dissynchrony a clue to deception? Insights from automated analysis of nonverbal visual cues. IEEE transactions on cybernetics, 45(3), pp.492-506.

[10] S. Lu, G. Tsechpenakis, D. N. Metaxas, M. L. Jensen, and J. Kruse, “Blob analysis of the head and hands: A method for deception detection,” in Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS ’05), 2005, pp. 1–10.

[11] L. Caso, F. Maricchiolo, M. Bonaiuto, A. Vrij, and S. Mann, “The impact of deception and suspicion on different hand movements,” J. Nonverbal Behav., vol. 30, no. 1, pp. 1–19, 2006.

[12] D. Cohen, G. Beattie, and H. Shovelton, “Nonverbal indicators of deception: How iconic gestures reveal thoughts that cannot be suppressed,” Semiotica, vol. 2010, no. 182, pp. 133–174, 2010.

[13] V. Pérez-Rosas, M. Abouelenien, R. Mihalcea, and M. Burzo, “Deception Detection using Real-life Trial Data,” IEEE Int. Conf. Multimodal Interact., pp. 59–66, 2015.

[14] L. ten Brinke and S. Porter, “Cry me a river: Identifying the behavioral consequences of extremely high-stakes interpersonal deception,” Law Hum. Behav., vol. 36, no. 6, pp. 469–477, 2012.

[15] P. Ekman and W. V Friesen, “The repertoire of nonverbal behavior: Categories, origins, usage, and coding,” Semiotica, vol. 1. pp. 49–98, 1969.

[16] Hoque, M.E., McDuff, D.J. and Picard, R.W., 2012. Exploring temporal patterns in classifying frustrated and delighted smiles. IEEE Transactions on Affective Computing, 3(3), pp.323-334.

[17] Li, R., Curhan, J. and Hoque, M.E., 2015, May. Predicting video-conferencing conversation outcomes based on modeling facial expression synchronization. In Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on (Vol. 1, pp. 1-6). IEEE.

[18] Jakobsson, Christer. "Peer-to-peer communication in web browsers using WebRTC." (2015).

[19] T. Baltrusaitis et al., “OpenFace: An open source facial behavior analysis toolkit,” in 2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016, 2016.

[20] Hutto, C.J. and Gilbert, E., 2014, May. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Eighth international AAAI conference on weblogs and social media.

[21] Pennebaker, J.W., Francis, M.E. and Booth, R.J., 2001. Linguistic inquiry and word count: LIWC 2001. Mahway: Lawrence Erlbaum Associates, 71(2001), p.2001.