

A Multimodal Dataset for Deception Detection

Verónica Pérez-Rosas¹, Rada Mihalcea², Alexis Narvaez³, Mihai Burzo⁴

¹Computer Science and Engineering, University of North Texas

²Computer Science and Electrical Engineering, University of Michigan

³Mechanical Engineering, University of North Texas

⁴Computer Science, Engineering, and Physics, University of Michigan - Flint

veronicaperezrosas@my.unt.com, mihalcea@umich.edu, AlexisNarvaez@my.unt.edu, mburzo@umich.edu

Abstract

This paper presents the construction of a multimodal dataset for deception detection, including physiological, thermal, and visual responses of human subjects under three deceptive scenarios. We present the experimental protocol, as well as the data acquisition process. To evaluate the usefulness of the dataset for the task of deception detection, we present a statistical analysis of the physiological and thermal modalities associated with the deceptive and truthful conditions. Initial results show that physiological and thermal responses can differentiate between deceptive and truthful states.

Keywords: deception detection, multimodal processing, corpus construction

1. Introduction

The act of deception can be defined as the action of intentionally causing another person to believe something that is known to be false by providing evidence to support that false belief (Mahon, 2007). Deception is present in our everyday life in the form of lies, misrepresentations, and omissions.

Identifying deceptive behavior is an open problem for the research community and has been studied under several contexts and perspectives. Research work has been extensively done in the psychology, security, and criminology areas. As a result, numerous clues have been identified as possible indicators of deceit. For example, physiological responses such as heart rate and skin responses; linguistic patterns such as self-references and cognitive words; facial behaviors such as blinking or smiling; body postures; and speech fluctuations.

These studies have been typically conducted over written deceptive and non-deceptive statements, one-to-one interviews, or by observing people behavior while being under a specific stimuli. Data acquisition in such cases was mainly focused on the clues to be analyzed and consists of either text statements, audio/visual recordings, or physiological sensor measurements. However, given the potential benefits of using multiple modalities at the same time to measure deceptive behaviors, it would be desirable to perform the data acquisition simultaneously over multiple data channels.

This paper describes the construction of a multimodal deception dataset. We use three deceptive scenarios to elicit deceptive and truthful responses. Data is collected in a multimodal setting where we acquire video/audio, thermal, and physiological recordings of the participants. Our initial results show that physiological and thermal responses can effectively differentiate between deceptive and truthful states.

2. Previous work

Deception on written content has been studied for a wide range of applications such as spam mail identification, anal-

ysis of public profiles in online websites, blog content and opinions (Hao et al., 2011). One common strategy is to analyze messages from deceivers and true tellers in order to find deception indicators. Linguistic clues such as self references or positive and negative words have been used to profile true tellers from liars (Newman et al., 2003). Other work has focused in identifying linguistic constructs such as the number of words, sentences, self references, affect, spatial and temporal information associated with deceptive content (Qin et al., 2005).

Research done towards deceit detection has also considered the measurement and analysis of physiological responses such as heart rate, skin conductance, and body temperatures (Frank et al., 2012). For instance, skin responses have been studied for the detection of criminal intent (Ewout H. Meijer and Merckelbach, 2010), whereas temperature profiles of per orbital and forehead areas have been analyzed to discriminate between deceptive and non deceptive behavior of subjects while being under a specific stimulus (Zhu et al., 2007).

There have also been several efforts to identify deceit using visual and acoustic clues. Experiments consisting of the elicitation of genuine, masked, simulated and neutralized emotions have been done to study involuntary emotional leakage in the form of micro expressions (Porter et al., 2012). Facial behaviors have been also evaluated by using the facial action units (AUs) system to identify deceptive responses (Ekman, 1993).

Finally, researchers have also focused instead on analyzing verbal behaviors exhibited by people while deceiving (Howard and Kirchhübel, 2011; Vrij et al., 2010). Speaking rate, energy, pitch, range as well as the identification of salient topics have been found useful to distinguish between deceptive and non-deceptive speech (Hirschberg et al., 2005).

3. Methodology

This section describes the process followed to create a multimodal dataset for deception detection. For this purpose,

we performed a set of experiments where subjects were asked to elicit deceptive or truthful responses according to three different scenarios, namely mock crime, best friend, and abortion.

Our choice of topics was motivated by previous research where linguistic behaviors have been studied under similar conditions and showed important differences between truth tellers and liars. However, unlike previous studies, we are incorporating visual, acoustic, physiological and thermal modalities to better sense differences between deceptive and truthful responses.

3.1. Materials

For our recordings we used a thermal camera FLIR Thermovision A40 with a resolution of 340x240 and a frame rate of 60 frames per second, and two regular web cameras: Quick Cam Orbit AF Logitech and HD webcam C525, each with a maximum resolution of 980x720 pixels and a frame rate of 30 frames per second. For physiological data acquisition, we used four different sensors: blood volume pulse (BVP sensor), skin conductance (SC sensor), skin temperature (T sensor), and abdominal respiration (BR sensor).

3.2. Participants

30 graduate and undergraduate students participated in the experiments. The sample consisted of 5 female and 25 male participants, all expressing themselves in English, from varied ethnical backgrounds (Asian, African-American, Caucasian, and Hispanic), with ages ranging between 22 and 38 years.

3.3. Procedure

We first presented each participant with an overview of the experimental procedure and hardware settings. We instructed them to respond either truthfully or deceptively, depending on the scenario being run. Also, two different stations were identified; one was a private area in the lab and the other was the experimental system station. Before each recording session, each participant was asked to sit at the recording station. Then, the participant was connected to the physiological sensors and both the thermal camera and the web cameras were adjusted according to each user height.

Participants were also asked to avoid any excessive movements with their head or hands because of the equipment motion sensitivity. The aim of the movement restrictions was to obtain high quality data from the cameras and the physiological sensors. This is particularly important for the temperature and the skin conductance measurements, since they are obtained using wired sensors that need to be in permanent contact with the skin. An alternative to this restrictive setting would be to use wireless sensors, which are non-invasive, provide acceptable accuracy, and allow mobility.

3.3.1. Scenario 1. Mock crime

In this experiment, participants were asked to be either deceptive or non deceptive in a mock crime scenario where they presumably stole a \$20 dollar bill. Each participant was told that in order to complete the task successfully, they

must not admit to the examiner that they have seen or taken the bill. Therefore, participants who did not have the bill were telling the truth while those who had the bill were lying. Experiment instructions for each participant were as follows:

1. Go to a private area of the lab and look for a hidden envelope, and take (presumably) a \$20 bill from this area.
2. Return to the recording station for a one-on-one interview.

Before the interview, the physiological sensors were calibrated for one minute in order to obtain an accurate sampling of the heart and breathing rates. Next, the examiner stated the following instructions: "Please listen and respond to each question truthfully. Please answer clearly and loudly. Please try not to move your head and shoulders during the interview, so that the sensors will not be affected by your movements.". The interview is as follows:

1. Are the lights on in this room?
2. Regarding that missing bill, do you intend to answer truthfully each question about that?
3. Prior to 2012, did you ever lie to someone who trusted you?
4. Did you take that bill?
5. Did you ever lie to keep out of trouble?
6. Did you take the bill from the private area of the lab?
7. Prior to this year, did you ever lie for personal gain?
8. What was inside the white envelope?
9. Please describe step by step, in as much detail as you can, what you did while you were behind the white board. Please aim at a clear description of about 2-3 minutes.
10. Do you know where that missing bill is now?

3.3.2. Scenario 2. Best Friend

In this experiment participants were asked to provide both a true description of their best friend, as well a deceptive description about a person they cannot stand. Each participant was asked to speak freely for about 2-3 minutes. The experiment session consisted of two independent recordings for each case, when a participant was either telling the truth or lying. Before each recording the examiner stated the following instructions:

1. First, think about your best friend. Talk about your friendship with him or her, mentioning the reasons for which you are such good friends, include anecdotes or anything that seems to you relevant to describe your relationship and what keeps you together. Thus, in this description you will have to tell the truth about what you feel about your best friend. Try to be clear, detailed and sincere. Please aim at a description of about 2-3 minutes.
2. Now think about a person you cannot stand, and talk about this person, describing this person as though he/she were your best friend. That is, in this description you will have to lie about what you actually feel about this person. Once again, please try to be as detailed as possible. Please aim at a description of about 2 - 3 minutes.

3.3.3. Scenario 3. Abortion

In this experiment the participants were asked to provide a truthful and a deceptive opinion about their feelings regarding abortion. Participants were asked to imagine a scenario where they took part in a debate on abortion in which they had an available time slot of 2-3 minutes to express their opinions.

The experiment session consisted of two independent recordings for each case, when the participant was either telling the truth or lying. In the first part of the experiment, the participant had to defend his or her point of view regarding abortion. In the second part of the experiment, the participant was asked to give a convincing statement defending the point of view opposite to his or her beliefs; in other words, the participant was asked to lie about what he or she really thinks about abortion. Before each recording the examiner repeated the following instructions:

1. Prepare a brief speech about your real opinion on abortion. Thus, in this speech you will have to tell the truth about what you believe about abortion. Remember that you will remain anonymous (your name will not be recorded), so please be truly honest in expressing your opinion. Try and explain the reasons that push you to take that position in the clearest, most detailed and most sincere way possible. Please aim at a speech of about 2-3 minutes.
2. Prepare a brief speech expressing the opposite of your opinion on abortion. That is, in this speech you will have to lie about what you actually believe about abortion. Once again, please try to be as detailed as possible. Please aim at a speech of about 2-3 minutes.

3.3.4. Thermal and Video Recordings

During each recording session two videos were obtained using the cameras described in section 3.1. The first camera was used to record a face close up whereas the second camera recorded the upper body, including the head, neck and arms. To acquire the thermal data we used the Flir ThermoCam Researcher software; in the case of the web cameras, we used the Logitech recording software. Two computers were used to control the signals and videos obtained; two experimenters were needed to control the data acquisition process and to conduct the interviews.

3.3.5. Physiological Measurements

Four sensors were attached to the non-dominant hand of the participants. Two skin conductance electrodes were placed on the second and third fingers whereas the skin temperature and blood volume blood volume sensors were placed at the thumb and index fingers respectively. The respiration sensor was placed comfortably around the thoracic region. The output of each sensor was obtained from a multimodal encoder connected to the main computer using an USB interface device. We recorded the combined output with the Biograph Infinity Physiology suite, which allowed us to visualize and control the data acquisition process.

4. Initial Experiments

4.1. Multimodal Features

After the data collection, we obtained a total of 30 observations for each scenario, including their corresponding visual, thermal, and physiological recordings. Table 1 shows the data distribution over the deceptive and truthful conditions. The raw data was then processed to obtain the following features to represent each modality.

Facial micro expressions: We automatically extracted 30 action units corresponding to muscle movements of eyes, eyebrows, nose, lips and chin. Additionally, we extracted smile intensity estimates as well as head pose orientation. Each feature was measured at frame level using CERT (Littlewort et al., 2011), and then averaged to obtain a single feature vector for each video recording.

Acoustic variations in speech: We extracted prosody, energy, voicing probabilities, spectrum, and cepstral coefficients to represent variations in speech. We use OpenEar(Schuller, 2009), an open source software for acoustic feature extraction. Overall, we obtained a set of 28 acoustic features. Each feature was computed using a frame sampling of 25ms and normalized using z-standardization. Since each feature was obtained at the frame level we averaged each feature value over all the frames in the audio recording to obtain a single feature vector.

Physiological responses: We obtained physiological features by processing the raw signal from each sensor. We used a commercial software called Biograph Infinity Physiology suite¹ to obtain physiological assessments for temperature, heart rate, blood volume pulse, skin conductance, and respiration. We obtained raw signal measurements as well as statistical descriptors for each physiological response, including maximum and minimum values, means, standard deviations, and mean amplitudes (epochs).

Thermal response: From the thermal recording, we obtained two different sets of features, one corresponding to face temperatures and the second one corresponding to the entire frame. Each set consisted of raw measurements of minimum and maximum temperatures, as well as statistical descriptors such as means and standard deviations. Each feature was measured at frame level and then averaged over the number of thermal frame samples.

4.2. Exploratory Analysis

As a preliminary analysis to identify relations between single modality features and deceptive behavior, we analyzed the percentile ranking among the deceptive and truthful groups. During this preliminary analysis, we also performed the percentile ranking analysis for visual and acoustic features, but they did not show a significant difference between groups.

¹<http://www.thoughttechnology.com/physuite.htm>

Scenario	Deceptive	Truthful
Mock crime	16	14
Best friend	30	30
Abortion	30	30

Table 1: Data distribution for deceptive and truthful conditions for three deceptive scenarios

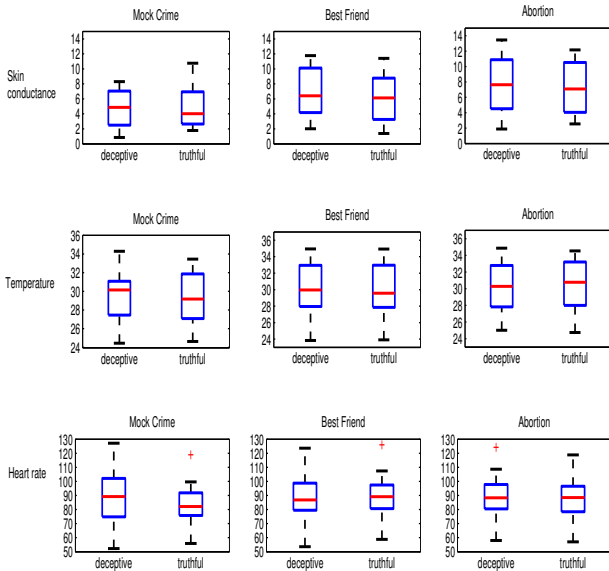


Figure 1: Percentile ranking analysis of physiological measurements: skin conductance, temperature, and heart rate

4.2.1. Physiological Responses

Figure 1 shows the box plots for the subjects' temperature, skin conductance, and heart rate responses during the mock crime, best friend and abortion scenarios.

As observed, for the mock crime scenario, physiological responses present important variations between the deceptive and truthful subjects. Raw skin conductance signal measurement is higher for the deceptive cases than for the truthful ones with an increment of two units. Temperatures also show a similar trend with a difference of about two degrees Celsius when people are being deceptive. Heart frequencies present higher differences between deceptive and truthful groups.

However, these trends are not applicable for the best friend and abortion scenarios where physiological responses are similar for both groups. As future work we are considering to further analyze these two scenarios and study the integration of additional modalities in order to find clearer differences between deceptive and truthful groups.

4.2.2. Thermal Responses

Figure 2 shows percentile ranking for the minimum and maximum temperature values collected with the thermal camera. From this graph, thermal measures show a clear differentiation between deceptive and truthful groups for the mock crime scenario. As before, differences are not clear for the best friend and the abortion scenarios.

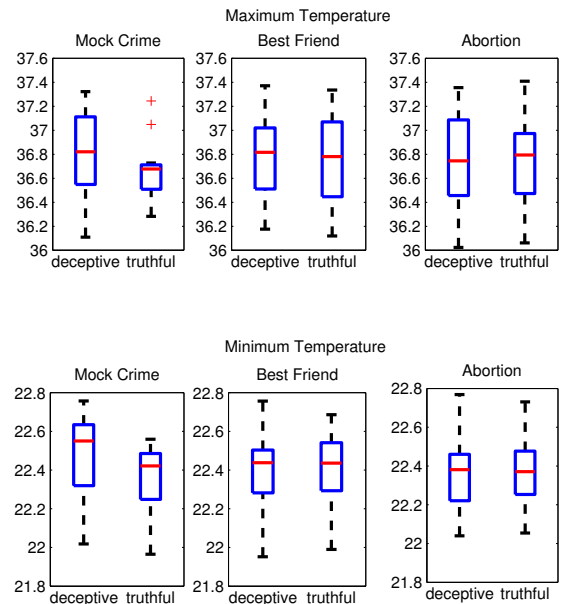


Figure 2: Percentile ranking analysis of thermal features

5. Conclusions

In this paper, we introduced a multimodal deception dataset, consisting of measurements of physiological, thermal, and visual responses recorded over three different deceptive scenarios. We presented our initial experiments including a statistical analysis of the differences between deceptive and truthful settings when using physiological and thermal responses. Our next step will be to perform multimodal experiments where we explore the role of each modality in deception detection.

The most important contribution of this paper is a new multimodal deception dataset, which includes accurate physiological measurements. These measurements can be used as ground truth for non-invasive approaches that use the visual and thermal data streams to estimate physiological measurements such as respiration and heart rate, thus making the process non-invasive and suitable for more realistic environments.

Acknowledgments

This material is based in part upon work supported by National Science Foundation award #1355633. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

6. References

- Ekman, P. (1993). Facial expression of emotion. *American Psychologist*, 48:384–392.
- Ewout H. Meijer, B. V. and Merckelbach, H. (2010). Detecting criminal intent with the concealed information test. *The Open Criminology Journal*, 3:44–47.
- Frank, M., Movellan, J., Bartlett, M., and Littleworth, G. (2012). RU-FACS-1 database, Machine Perception Laboratory, U.C. San Diego.

- Hao, P., Chen, X., Cheng, N., Chandramouli, R., and Subbalakshmi, K. P. (2011). Adaptive context modeling for deception detection in emails. In *Proceedings of the 7th international conference on Machine learning and data mining in pattern recognition, MLDM'11*, pages 458–468, Berlin, Heidelberg. Springer-Verlag.
- Hirschberg, J., Benus, S., Brenier, J. M., Enos, F., Friedman, S., Gilman, S., Gir, C., Graciarena, M., Kathol, A., and Michaelis, L. (2005). Distinguishing deceptive from non-deceptive speech. In *In Proceedings of Interspeech; 1/2 2005 - Eurospeech*, pages 1833–1836.
- Howard, D. M. and Kirchhübel, C. (2011). Acoustic correlates of deceptive speech: an exploratory study. In *Proceedings of the 9th international conference on Engineering psychology and cognitive ergonomics, EPCE'11*, pages 28–37, Berlin, Heidelberg. Springer-Verlag.
- Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., and Bartlett, M. (2011). The computer expression recognition toolbox (cert). In *Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 298 –305, march.
- Mahon, J. E. (2007). A definition of deceiving. *International Journal of Applied Philosophy*, 21(2):181–194.
- Newman, M., Pennebaker, J., Berry, D., and Richards, J. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, 29.
- Porter, S., ten Brinke, L., and Wallace, B. (2012). Secrets and lies: Involuntary leakage in deceptive facial expressions as a function of emotional intensity. *Journal of Nonverbal Behavior*, 36(1):23–37.
- Qin, T., Burgoon, J. K., Blair, J. P., and Nunamaker, J. F. (2005). Modality effects in deception detection and applications in automatic deception-detection. In *Proceedings of the 38th Hawaii International Conference on System Sciences*.
- Schuller, F. E. M. W. B. (2009). Openear - introducing the Munich open-source emotion and affect recognition toolkit. In *ACII*.
- Vrij, A., Granhag, P. A., and Porter, S. (2010). Pitfalls and Opportunities in Nonverbal and Verbal Lie Detection. *Psychological Science in the Public Interest*, 11(3):89–121, December.
- Zhu, Z., Tsiamyrtzis, P., and Pavlidis, I. (2007). Forehead thermal signature extraction in lie detection. *Conf Proc IEEE Eng Med Biol Soc*, 2007.