First Attempts in Deception Detection in HRI by using Thermal and RGB-D cameras
David-Octavian Iacob, Adriana Tapus

To cite this version:
David-Octavian Iacob, Adriana Tapus. First Attempts in Deception Detection in HRI by using Thermal and RGB-D cameras. RO-MAN 2018, Aug 2018, Nanjing, China. hal-01840122

HAL Id: hal-01840122
https://hal.archives-ouvertes.fr/hal-01840122
Submitted on 16 Jul 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
First Attempts in Deception Detection in HRI by using Thermal and RGB-D cameras

David-Octavian Iacob and Adriana Tapus

Abstract—A key factor of a successful Human-Robot Interaction (HRI) is the robot’s ability to understand the messages transmitted by a human user, especially their nonverbal and paraverbal components. Our work focuses on the robot’s ability to detect if a person is deceitful or honest, based on an evaluation of their physiological state and using only noninvasive techniques, such as thermal and RGB-D imaging. In particular, we currently study and evaluate several physiological manifestations during an experiment consisting of an interrogation conducted by a robot that follows a mock-up crime, then correlate them with the truth values of the participants’ answers. The first results are promising.

I. INTRODUCTION

Social robotics [1][2] focuses on the development of robots that are able to interact as naturally as possible with humans. For a Human-Robot Interaction (HRI) to be natural and efficient, the robots should, first of all, be able to understand the messages sent by their human interlocutors as fast and as accurately as a human being. In this work, we do not focus on the verbal aspects of HRI, but on the nonverbal ones. Without the ability to understand the nonverbal and paraverbal components of the communication, the robots cannot fully understand the messages sent by their interlocutors. One of the consequences of a robot’s inability to understand the nonverbal aspects of the communication is that humans could much more easily deceive a robot during an interaction than it would be to deceive another human in a similar human-human interaction.

Socially Assistive Robots (SAR) [3] focus on the social component of the interaction with humans in order to assist them. Their objective is to improve the quality of life of certain populations of users, such as people with neurological [4], motor [5], and/or cognitive impairments [6] in need of physical or psychological rehabilitation. Nonetheless, the latter might try to deceive the former the same way they might try to deceive a human assistant checking up on their progress (e.g., not taking their medication, not doing their physical therapy [7], not respecting certain interdictions or recommendations). Therefore, in order to improve the quality of the interaction, assistive robots need to be able to establish if the human users they interact with are deceitful or trustworthy, at least up to the extent a human assistant is capable of doing so in similar scenarios.

Our work focuses on the development of noninvasive techniques that would allow a robot to detect deception, based on a series of physiological manifestations. Some of the physiological parameters are nonverbal cues, such as the head position and orientation, the eye openness and position, while others describe the cardiovascular and respiratory response of the organism to the psychological state of the individual, such as the heart rate, respiratory rate, and facial skin temperature variations. These parameters are monitored using noninvasive sensors, such as RGB-D [8] and thermal cameras [9][10][11]. The observations are then correlated with a known psychological state of the human user (either deception or trustworthiness). In order to measure, evaluate, and compare these manifestations, we have designed an experiment where human participants are enticed to deceive their robot interlocutor in order to get a reward.

The rest of the paper is structured as follows: section II presents an overview of the techniques that can be used in order to detect deception, section III describes the experimental setup we have used for our experiment, section IV describes the methodology used for data analysis, section V discusses the results of our work, and finally section VI concludes our paper.

II. DECEPTION AND ITS PHYSIOLOGICAL MANIFESTATIONS

The development of the polygraph [12] has shown that several physiological parameters can be correlated with deception, with various degrees of accuracy. We can distinguish between invasive measurement techniques, used to measure parameters such as the galvanic skin response (GSR) (or other measurements of the electro-dermal activity) [12][13][14], heart rate and blood pressure [12][15], respiratory rate [12] or brain activity [16][17] and noninvasive measurement techniques, that can be used to monitor heart rate, using Fast Fourier Transform (FFT) video analysis [8][18], respiratory rate (FFT video analysis) [19][20], voice audio frequencies[21][22], facial expressions[22], or eye movements and blink rates[23][24][28].

The invasive physiological evaluation methods are, by far, the most accurate, since the measurement errors of the devices used in order to monitor these parameters are very low and allow a very accurate tracking of human physiological manifestations. According to [25], polygraphs can now offer a reliability of 81% to 91%. Moreover, many of them have been used for several decades by prosecutors,
investigators, and police forces as part of standardized lie detection techniques, therefore being considered the most advanced lie detection techniques available today. Nonetheless, even the best deception detection systems could provide false positive or false negative results, depending on each individual’s physiological response, as well as their ability to control their physiological manifestations.

The main problem with the invasive lie detection techniques is the fact that some of them cannot be used in everyday human-robot interactions or human-human interactions, since they involve complex, fixed devices that cannot be carried around, while minimally-invasive techniques still mean that the person being monitored may simply remove the sensors placed on their body in order to avoid any form of deception detection. The heart rate [26][27] and respiratory rate [27] can be evaluated using minimally invasive techniques, as well as the electro-dermal activity [28] (up to some extent, depending on the number and location of sensors). Blood pressure measurements [15] involve more encumbering sensors, while brain activity mapping [16][17] requires very complex and generally fixed, rooted systems.

Noninvasive lie detection has been already attempted in inter-human interaction scenarios [9][10][21][22][24] with various degrees of success. Moreover, as previously shown, a series of noninvasive physiological measurements techniques have already been developed and they could be used for noninvasive lie detection. Nonetheless, to the best of our knowledge, there are no other works focusing on noninvasive physiological measurement based lie detection in HRI.

Therefore, one of the main objectives of this research is to develop purely noninvasive lie detection techniques that can easily be integrated in HRI scenarios, without interfering with the fluidity of the interaction. Any such device and method should therefore use sensors that can easily be integrated in a robot (if they are not already part of some robotic platforms) and should use algorithms that could evaluate if the interlocutor is deceitful or not with a minimum delay using computation power that can be provided by regular portable computers (laptops or embedded systems such as Raspberry PI and similar systems).

III. EXPERIMENTAL SETUP

The experiment we have designed has two main parts, that take place in different rooms: the mock-up crime and the interrogation. After the experiment is over, the participants are asked to fill a post-experimental questionnaire as well as a BIG5 [29] personality test. This experiment has been designed in order to collect relevant data and then to create a model based on the correlations that are found.

A. Part 1: The mock-up crime

After the participant was informed of the experiment, they enter the first room alone, then they locked the door. On a table inside the room, there is a sheet of paper containing the detailed experimental instructions for this part, a chocolate bar, and five objects: a red marker, a computer mouse, a €20 bill, a book, and a mobile phone.

The participant is asked to choose 3 out of the 5 objects and hide them anywhere inside the room. Afterwards, the participant must take the chocolate bar, leave the room and not tell anyone, human or robot, what objects they have hidden inside the room, or otherwise they will not be able to keep the reward (the chocolate bar) at the end of the experiment. Afterwards, they are instructed to go straight to the interrogation room.

The room is secretly kept under video surveillance using a hidden wireless camera that allows the experimenter to know exactly what objects have been hidden by the participant before they start the second part of the experiment. This information is used in order to adapt the robot’s interrogation speech.

B. Part 2: The interrogation

The participant enters the interrogation room, where an interrogator (a Pepper robot) sits behind a table. No humans can be seen inside the room, as the experimenter is hidden behind some curtains and oversees the experiment. The participant is therefore not aware of the experimenter’s presence. The participant is invited by the robot to take a seat on the other side of the table, in front of the interrogator, who will then explain the interrogation procedure.

The interrogation is recorded with the RGB-D and thermal camera system and consists of four phases:

1) General questions, not related to the experiment (e.g., Are you sitting down right now?).
2) Questions concerning the first part of the experiment to which the participant has no reason to lie (none is related to the 3 objects they have hidden).
3) Questions concerning the first part of the experiment to which the participant is enticed to lie (all of them are related to the 3 objects they have hidden).
4) The same exact general questions, not related to the experiment, asked in the first phase of the interrogation.

The participant is told that they are free to choose whether they are honest or not throughout the interrogation. Therefore, the datasets obtained from each participant may have a variable number and distribution of truthful and dishonest
answers, but the participants are given the control concerning what elements of information they are willing to disclose or not. This makes the interrogation process more realistic, even though the data is harder to analyze.

The interrogation begins with 11 preliminary questions, which the participant has no particular reason to answer dishonestly. Their purpose is to establish a baseline for the physiological parameters that are being monitored. Then the actual interrogation begins, with 11 questions concerning details of the first part of the experiment, but none related to the manipulation of the three objects the participant has hidden. The participants may choose to lie anyway, either because they do not want to tell anything about any aspect of the first part of the experiment or because they are unsure what details they have to hide in order to keep the reward.

The third part of the interrogation consists of 11 questions related to the three objects that have been hidden by the participant. This is the part of the interrogation where the participants must absolutely hide the truth in order to keep their reward. After this interrogation phase is done, the same 11 preliminary questions that have been asked in the beginning of the interrogation are being asked again. The purpose of this repetition of the preliminary question phase is to study if and to what extent the participants’ physiological manifestations evolve back towards the baseline values. At the end of the experiment, the robot informs the participant if they can or cannot keep their reward. The participant is then invited to leave the room.

The interrogator (the Pepper robot) is remotely controlled by the experimenter supervising the interrogation behind the curtains. Several decisions have to be made based on the video footage provided by the hidden camera: which questions have to be included in the second and third part of the interrogation (based on the hidden objects) and whether the participant can or cannot keep their reward (based on the information revealed by the participant during the interrogation). Since the robot interrogator has no ability to detect lies so far, these decisions are taken by the human supervisor based on an assessment of the participant’s answers.

Another key aspect of the interrogation procedure is that instead of using Pepper’s default voice and text-to-speech, the questions are pre-recorded using the experimenter’s voice, in both English and French. The main reason for this is that the default voice of the Pepper robot resembles that of a child, making it unsuitable for such interrogation scenarios. The questions (and the audio recordings used) are exactly the same for all the participants, except, of course, for the names of the hidden objects, which vary depending on what each participant has decided to hide, and the intonation used was as neutral and easy to understand as possible.

C. The questionnaires

Right after the experiment is over, the participants are asked to fill up a post-experimental questionnaire. It consists of a series of questions asking for feedback concerning the experiment, using a 7-points Likert scale. The post-experimental questionnaire’s objective is to establish the perception of the participants with respect to several aspects:

- how much they wanted to keep the reward (as their desire to eat a chocolate bar might vary)
- what was their perceived level of stress during each phase of the interaction (as stress can have similar physiological manifestations as deception)
- whether they thought the robot interrogator knew the truth concerning what they have done in the first part of the experiment
- whether they thought the robot interrogator was able to detect if they told the truth and if they lied (since they are not told whether the robot has this ability or not)
- whether they considered the questions asked during the interrogation relevant in order to establish what they have done during the first part of the experiment.

The post-experimental questionnaires are either used to check if there is a certain bias in the experimental procedure (such as the reward not being enticing enough for the participants or the experimental procedure being too stressful) and also to correlate the physical manifestations of each participant to its perception of the experiment.

Another questionnaire that the participants are asked to fill up is the BIG5 personality test. For each of the five measured personality traits, the participants received a score from 1 to 5. In particular, we are interested in studying if their extroversion, their neuroticism, and their conscientiousness can be correlated with their physiological manifestations when trying to deceive the robot interrogator.

D. Participants

A number of 20 participants took part in this experiment. We have established, based on the post-experimental questionnaire as well as the BIG5 personality test, several subgroups, depending on each of the following parameters:

- gender: 13 male, 7 female
- perceived stress (interrogation): 12 low, 8 high
- desire to keep the reward: 4 low, 16 high
- extroversion: 8 introverts, 12 extroverts
- conscientiousness: 8 low, 12 high
- neuroticism: 9 low, 11 high

They were split based on the comparison of the value of each parameter with the average value (which is 3 in both the post-experimental questionnaire and the BIG5 personality test). A large majority of the participants, 17 out of 20, have a background in Technical Sciences, one in Theatre, one in Physics, and one in Medical Sciences.

IV. DATA ANALYSIS

A. Measurements

Two video cameras record the interrogation: an Asus Xtion RGB-D camera and an Optris Pi640 thermal camera. They both provide a 640x480 output at 30 Hz and are connected to the data analysis computer used by the experimenter to supervise the experiment and to remotely control the robot.
The video streams are being acquired using two ROS packages and then processed using a program developed in C++ using ROS and OpenCV. The software uses the CLM Face Tracker [30] to track 66 facial landmark points in the RGB-D frame, then a coordinate conversion function that transforms the RGB-D frame coordinates to thermal frame coordinates, using a calibration procedure based on the ROS camera_calibration package. The application also monitors the video stream provided by the hidden camera in the first experimental room, allowing the experimenter to configure the robot’s behaviour (which questions to ask) before the participant starts the interrogation.

Based on the facial landmark points detected using the CLM Face Tracker, several regions of interest (ROI) are being identified in the RGB frame, then their coordinates are being converted to thermal frame coordinates. The main regions of interest and the respective evaluated parameters are the following:

- **the facial area**: position, orientation, green colour component, depth (distance), temperature
- **the ocular and periocular areas**: relative position, eye openness, relative temperature
- **the nose area**: relative temperature

The average values of each variable are being computed for each region of interest, while the relative positions and temperatures are computed with respect to the average values computed for the entire face. Each value, once computed, is logged to an individual stamped CSV (Comma Separated Variable) file, that is later used for post-experimental processing and analysis.

**B. Analysis procedure**

The logged data is analyzed post-experimentally, using Scilab / MATLAB scripts that process the CSV log files. Manual annotations register the key moments of the interrogation (beginning and end of each interrogation block), as well as the number of true and false answers for each interval. The annotations are used for data segmentation, so the evolution of each parameter can be correlated to the honesty of the participants.

For a given participant and evaluated parameter, we define 4 analysis intervals:

- preliminary questions (first run)
- experiment-related questions: expected truths
- experiment-related questions: expected lies
- preliminary questions (second run)

If a participant has given more than 50% honest answers in a given interval, we consider that they have been honest during that interval and dishonest otherwise. If an answer is not clearly honest nor dishonest, it is marked as inconclusive and not taken into account. Therefore, for each of these four intervals we associate a “honesty” parameter that is either 1 or 0 depending on the percentage of truthful answers. The associated “honesty” parameter is 0 if the percentage of truthful answers is strictly lower than 50% and 1 otherwise.

All the recorded data was initially annotated by one of the authors. Afterwards, the data recordings of 3 random participants has been chosen for annotation by a second rater. The Cohen’s kappa coefficient for inter-rater agreement had a value of $\kappa = 0.921$, which corresponds to almost perfect agreement.

**V. RESULTS**

**A. Statistical correlations**

Based on the evaluation of the logged parameters, the physiological parameters we monitored are:

- estimated heart rate (facial green FFT)
- estimated respiratory rate (nasal temperature FFT)
- interocular and nasal relative temperature
- eye vertical relative position
- face distance with respect to the cameras

Our first strategy was to analyse the data for each question/answer pair and correlate it with the truth value of each given answer. This approach was problematic for several reasons. First of all, due to the limited sampling rate of the video cameras (30 Hz), the FFT analysis used for the heart rate and respiratory rate detection algorithms needed windows of 30-60 seconds in order to have an usable frequency resolution:
1 beat per minute (BPM) for 60 seconds and 2 BPM for 30 seconds. On the other hand, some physiological manifestations, such as temperature variations due to vasodilation and vasoconstriction cannot be measured instantly, due to the body’s response times and the skin’s thermal inertia, which also implies a need for larger measurement windows.

Among all the evaluated physiological parameters, the only one who has shown a statistical correlation was the estimated heart rate variability and only on three experimental subgroups: males, females, and participants with high neuroticism.

The heart rate was estimated based on a weighted mean of the most significant 5 frequency peaks of the FFT analysis in the 1...3 Hz frequency domain (corresponding to 60-180 BPM), since we assumed that the participants did not have a constant heart rate during the reference intervals and that the data measurements were prone to errors. The following graphs show the heart rate variability in Hz with respect to the first measurements interval (preliminary questions) when the participants have mostly lied (column 0) or mostly told the truth (column 1). For each of these three experimental subgroups we have proven that there is a statistically significant difference in the heart rate variability with respect to the preliminary questions interval when saying the truth compared to when they were lying.

The 7 female participants have shown an estimated mean heart rate variability of -0.004 Hz (-0.24 BPM) when lying and of 0.2 Hz (+12 BPM) when telling the truth, with \( F(1,19)=5.87 \) and \( p=0.0256 \). The data was normally distributed and analyzed with the ANOVA test.

The 11 participants with high neuroticism have shown an estimated average heart rate variability of -0.021 Hz (-1.26 BPM) when lying and of +0.167 Hz (+10.02 BPM) when telling the truth, with \( F(1,31)=5.28 \) and \( p=0.028 \). The data was normally distributed and was analyzed with the Kruskal-Wallis test.

The 13 male participants have shown an estimated average heart rate variability of -0.185 Hz (-11.1 BPM) when lying and of +0.01 Hz (+0.6 BPM) when telling the truth, with \( \chi^2(1,37)=5.73 \) and \( p=0.0166 \). The data was not normally distributed and was analyzed with the Kruskal-Wallis test.

The other variables that were analyzed, both on the entire group of 20 participants and on each individual subgroup have not shown any relevant statistical correlations according to standardized statistical tests. This can be caused by the low number of participants that have undertaken the experiment so far.

**B. Questionnaire-based evaluation**

The answers of the participants to the post-experimental questionnaire have been evaluated in order to establish whether the experiment was well designed or not. The participants’ answers are on a Likert scale from 1 to 7, where 1 corresponds to “Strongly disagree”, 7 corresponds to “Strongly agree” and 3 to “Neither agree nor disagree”.

1) *Did you want to keep the chocolate bar?*

   Average: 5.50 (standard deviation of ±2.01)
2) Do you think the interrogator knew everything that you did inside the first room?
   Average: 5.05 (standard deviation of ±2.06)
3) Do you think the interrogator was able to detect when you were telling the truth and when you were not?
   Average: 3.96 (standard deviation of ±1.73)
4) Did you find the first part of the experiment stressful?
   Average: 1.75 (standard deviation of ±1.29)
5) Did you find the interrogation stressful?
   Average: 3.85 (standard deviation of ±2.00)
6) Did you feel threatened by the interrogator at any moment?
   Average: 2.8 (standard deviation of ±1.54)

We can conclude that the reward was considered generally desirable by the participants, in spite of its low cost and of each participant’s personal preference for chocolate. The participants were also generally convinced that the robot interrogator knew for a fact what objects they have hidden during the first part of the experiment, even though they were not as convinced that the robot had actual lie detection abilities.

Their evaluation of the stress level was also important. The first part of the experiment was considered totally not stressful, while the perceived level of stress during the interrogation was slightly above average. Moreover, the perceived level of threat from the robot interrogator was also slightly below average. This means that the interrogation could have been designed to be less stressful and/or threatening, in order to further reduce the risk of mistaking the physiological manifestations correlated to stress with those correlated to deception.

VI. CONCLUSIONS

The purpose of this research was to design an HRI scenario during which humans would be enticed to lie to a robot, therefore allowing us to evaluate their nonverbal and physiological manifestations using only noninvasive techniques and then to correlate the measurements with the honesty or dishonesty of their answers.

We have found statistical correlation between the truth value of the participants’ answers for each of the measurement intervals and the estimated heart rate variability with respect to the first interval (preliminary questions) in three of the experimental subgroups (males, females, and participants with high neuroticism). In each of the three subgroups, the estimated heart rate was significantly higher when participants were mostly telling the truth than when they were mostly lying to the robot interrogator.

Nonetheless, the lower accuracy of the noninvasive measurements had a negative effect on our results meant that among all the monitored parameters the only one that was successfully statistically correlated with deception was the estimated heart rate variability. This was mostly due to the measurement errors induced by the measurement techniques and the hardware limitations, as well as by the fact that we did not restrict the natural movements of the participants in any way. This experiment can be considered in the wild, no restrictions, and no constraints were imposed to the participants.

In future research, we plan focusing on various ways of improving the quality of the data acquisition so that we are able to accurately monitor the other physiological manifestations that we tried to analyze during this experiment. Moreover, other noninvasive techniques will be integrated in order to evaluate more parameters that have been proven to be correlated with deception.

Moreover, we will also study whether there are any differences between the manifestations associated with deception when the participants lie to a robot or another human in identical interaction scenarios.

ACKNOWLEDGMENT

This research has been financed by the French Ministère de l’Enseignement supérieur, de la Recherche et de l’Innovation, through the INTERFACES Doctoral School. We would equally like to thank all the participants that took part in our experiments, as well as to Denis Dulgeriu for his help with the annotation process.

REFERENCES