Lie detection in human-robot interactions using noninvasive and minimally-invasive devices and sensors

David-Octavian Iacob

INSTITUT POLYTECHNIQUE DE PARIS

NNT : 2019IPPAE004

Thèse de doctorat

ENSTA

IP PARIS

Détection du mensonge dans le cadre des interactions homme-robot à l'aide de capteurs et dispositifs non invasifs et mini invasifs

Thèse de doctorat de l'Institut Polytechnique de Paris
préparée à l'Ecole Nationale Supérieure de Techniques Avancées

École doctorale n°626 Ecole Doctorale IP Paris (ED IP Paris)
Spécialité de doctorat : Informatique

Thèse présentée et soutenue à Palaiseau, le 18/11/2019, par

DAVID-OCTAVIAN IACOB

Composition du Jury :

David FILLIAT Professeur, ENSTA Paris (U2IS) - France	Président
Mehdi AMMI Professeur, Université Paris 8	Rapporteur
Silvia ROSSI Professeur, Université "Federico II" de Naples (DIETI) - Italie	Rapporteur
Rachid ALAMI Professeur, LAAS-CNRS	Examinateur
Adriana TAPUS Professeur, ENSTA Paris (U2IS) - France	Directeur de thèse

Declaration of Authorship

I, David-Octavian IACOB, declare that this thesis titled, "Lie detection in human-robot interactions using noninvasive and minimally-invasive devices and sensors" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

Acknowledgements

First and foremost, I would like to thank my supervisor, Adriana Tapus, for guiding and supporting me, not only in the past three years that I've dedicated to this thesis, but ever since I first arrived at ENSTA Paris, in the summer of 2015. My journey followed the roads that brought me here today thanks mainly to her help, advice and patience, and I will be forever grateful for it.

Then, I would like to thank all the teachers, researchers, and interns of the U2IS laboratory, for their contribution to my work, for being great brainstorming partners, for taking part in my experiments and for all the great memories we have shared together. I am also grateful to all my former teachers, of the "Emil Racoviță" National College of Cluj-Napoca, the Technical University of Cluj-Napoca, Ecole Polytechnique, and ENSTA Paris, for shaping me into who I am today and for giving me the tools I needed in order to pursue my goals. On the same note, I am also grateful to all the people who started and nurtured the fire that is my passion for robotics.

Finally, I would like to thank all the participants to my experiments for accepting to contribute to my work, as well as all my friends, who supported me through thick and thin, helped me stay on the course set towards my dreams and goals, sometimes from hundreds or thousands of kilometres away. Last, but definitely not least, I would like to thank my parents, for encouraging me to pursue this PhD thesis and for supporting me unconditionally.

Contents

Declaration of Authorship iii

Acknowledgements vii

1 Introduction 1
- 1.1 Research context and motivation . . . 1
  - 1.1.1 Why detect deception in HRI? . . . 2
  - 1.1.2 Challenges of in-the-wild HRI . . . 3
- 1.2 Thesis overview . . . 4

2 Physiology of deception 5
- 2.1 Introduction . . . 5
- 2.2 Approaches to detecting deception . . . 5
  - 2.2.1 Polygraph-based methods . . . 6
  - 2.2.2 Cognitive load based methods . . . 7
- 2.3 Manifestations associated with deception . . . 8
  - 2.3.1 Physiological manifestations . . . 9
  - 2.3.2 Behavioural manifestations . . . 12
- 2.4 Measurement devices and techniques . . . 15
  - 2.4.1 Invasive measurement devices . . . 16
  - 2.4.2 Noninvasive measurement devices . . . 17
  - 2.4.3 Minimally-invasive measurement devices . . . 19
- 2.5 Applying theories of inter-human interactions to HRI . . . 21
- 2.6 Conclusions . . . 21

3 Experimental methodology and resources 23
- 3.1 Introduction . . . 23
- 3.2 Experimental design guidelines . . . 23
- 3.3 Robotic platforms and sensors . . . 25
  - 3.3.1 Robotic platforms . . . 25
  - 3.3.2 Sensors and measurement devices . . . 26
- 3.4 Experimental monitoring and analysis software . . . 30
  - 3.4.1 Experimental software application . . . 31
  - 3.4.2 Post-experimental data analysis . . . 35
- 3.5 Personality profiling . . . 40
- 3.6 Summary . . . 42
- 3.7 My contribution . . . 42

4 Detecting deception using only RGB-D and thermal cameras 43
- 4.1 Introduction . . . 43
- 4.2 Experimental setup: mock-up crime and interrogation . . . 43
  - 4.2.1 The mock-up crime . . . 44
  - 4.2.2 The interrogation . . . 44

4.2.3 Annotating the experimental data . . . 46
4.2.4 Participants . . . 47
4.3 Detecting deception using RGB-D cameras . . . 48
4.3.1 Tracking the interaction distance . . . 48
4.3.2 Eye openness . . . 51
4.3.3 Estimated heart rate . . . 52
4.4 Thermal image analysis . . . 55
4.4.1 Tracking facial ROIs . . . 56
4.4.2 Facial area temperature variations . . . 59
4.4.3 Respiratory rate estimation . . . 60
4.5 Post-experimental questionnaires . . . 61
4.5.1 Objectives and questions . . . 61
4.5.2 Analysis of results . . . 62
4.6 Summary . . . 62
4.6.1 Results . . . 63
4.6.2 Limitations and future developments . . . 63
4.7 My contribution . . . 64

5 Using wearable sensors in conjunction with an RGB camera 65
5.1 Introduction . . . 65
5.2 Experimental setup: the card guessing game scenario . . . 65
5.2.1 The decks of cards . . . 66
5.2.2 The game rules . . . 66
5.2.3 The experiment . . . 67
5.2.4 Annotating the experimental data . . . 68
5.2.5 Participants . . . 68
5.3 Detecting deception using RGB cameras . . . 69
5.3.1 Tracking the head position and orientation . . . 70
5.3.2 Eye openness . . . 74
5.4 Detecting deception using the wireless armband . . . 75
5.4.1 Heart rate . . . 75
Measurement technique . . . 75
5.4.2 GSR - Skin conductance . . . 77
5.5 Post-experimental and post-round questionnaires . . . 78
5.5.1 Post-experimental questionnaire . . . 78
5.5.2 Post-round questionnaire . . . 80
5.6 Summary . . . 81
5.6.1 Results . . . 81
5.6.2 Discussion . . . 81
5.7 My contribution . . . 82

6 Comparing human-robot and human-human interactions 83
6.1 Introduction . . . 83
6.2 Experimental setup and evaluated parameters . . . 83
6.2.1 Experimental setup review: the card guessing game . . . 84
6.2.2 Evaluated parameters and analysis methods . . . 84
6.3 Results . . . 85
6.3.1 Discussion . . . 86
6.4 Post-experimental questionnaires . . . 88
6.5 Summary . . . 89
6.6 My contribution . . . 90

7 Detecting deception using audio analysis **91**
7.1 Introduction . . . 91
7.2 Experimental setup and data analysis . . . 92
7.2.1 Audio analysis procedure: extracting the response times . . . 92
7.3 Results . . . 94
7.4 Summary . . . 95
7.5 My contribution . . . 96

8 Overview of results **97**
8.1 Introduction . . . 97
8.2 Physiological manifestations associated to deception . . . 97
8.2.1 Heart rate . . . 98
8.2.2 Skin conductance . . . 98
8.2.3 Head position and orientation . . . 99
8.2.4 Response times to questions . . . 99
8.3 HRI vs. inter-human interactions . . . 100
8.3.1 Skin conductance . . . 100
8.3.2 Eye openness . . . 100
8.3.3 Head position and orientation . . . 100
8.4 Post-experimental questionnaire evaluation . . . 101
8.5 Publications . . . 102

9 Conclusions **103**
9.1 Introduction . . . 103
9.2 The significance of our results . . . 103
9.3 Perspectives . . . 104

Academic CV **107**

References **109**

A The BIG5 Personality Test **117**

B The short, revised version of the Eysenck Personality Questionnaire **119**

C Question tree used for the card guessing game **121**

List of Figures

3.1 Meka, Pepper, and TIAGo robots from the Autonomous Systems and Robotics lab/U2IS . . . 26
3.2 The Asus Xtion RGB-D camera . . . 27
3.3 The Logitech C920 Webcam . . . 28
3.4 The Optris PI640 thermal camera and an example of a thermal image representation . . . 28
3.5 The wireless sensor armband . . . 29
3.6 The circuits inside the wireless sensor armband . . . 30
3.7 The final version of our experimental GUI . . . 32
3.8 The evolution and the processing of a participant's HR . . . 36
3.9 FFT analysis of the facial green colour component . . . 38

4.1 The mock-up crime room: the table, the objects and the hidden camera 44
4.2 The interrogation room setup and the Pepper robot . . . 45
4.3 CLM Face Tracker landmarks . . . 49
4.4 Facial ROIs and landmarks . . . 50
4.5 Female participants: heart rate variability . . . 54
4.6 Male participants: heart rate variability . . . 55
4.7 Participants with high neuroticism: heart rate variability . . . 55
4.8 The LED matrix used for the camera calibration procedure . . . 57
4.9 The interface of the LED matrix detection software . . . 58
4.10 Using the OpenCV camera calibration package . . . 58
4.11 RGB and thermal face tracking - coordinate conversion . . . 59

5.1 The two decks of cards . . . 66
5.2 Playing the card game with a robot game partner . . . 67
5.3 DLib Face Tracker landmark points . . . 71
5.4 Relative face y coordinate average (robot game partner) . . . 73
5.5 Relative heart rate linear interpolation slope (robot game partner) . . . 76
5.6 Perceived difficulty and stress level evolution . . . 80

7.1 Display of the response time extraction algorithm . . . 93
7.2 The response time to questions (seconds) . . . 94
7.3 The distribution of the response time to questions (seconds) . . . 95

List of Tables

2.1 A review of manifestations associated to deception 15

4.1 BIG5 personality test analysis 48
4.2 The results of the mock-up crime & interrogation post-experimental questionnaire 62

5.1 The 32 items displayed on the picture cards 66
5.2 Eysenck personality questionnaire analysis 69
5.3 Head position and orientation: overall results 72
5.4 Head height relative linear interpolation slope: subgroups analysis . . 73
5.5 The results of the card guessing game post-experimental questionnaire 79
5.6 Overall results 81

6.1 Analysed measures vs. nature of the game partner 85
6.2 Analysed measures: comparison between the honesty condition and the game partner condition 87
6.3 Post-experimental questionnaire answers: average and distribution . . 88

8.1 Results summary: heart rate 98
8.2 Results summary: skin conductance 98
8.3 Results summary: head position and orientation 99
8.4 Results summary: response times 99
8.5 Results summary: skin conductance 100
8.6 Results summary: eye openness 100
8.7 Results summary: head position and orientation 101

C.1 Card guessing question tree: phrasing A 121
C.2 Card guessing question tree: phrasing B 122

List of Abbreviations

ANS	**A**utonomous **N**ervous **S**ystem
BPM	**B**eats **P**er **M**inute
CSV	**C**omma **S**eparated **V**ariables
GSR	**G**alvanic **S**kin **R**esponse
HR	**H**eart **R**ate
HRI	**H**uman **R**obot **I**nteraction(s)
PPG	**P**hoto**p**lethysmo**g**raphy
ROI	**R**egion(s) **O**f **I**nterest
ROS	**R**obot **O**perating **S**ystem
SC	**S**kin **C**onductance
SR	**S**kin **R**esistance

Chapter 1

Introduction

Robots, unlike many other technological achievements of recent times, were at first the result of the fantasy of writers. The first time this concept was defined was in 1920, in Karel Čapek's science fiction play entitled R.U.R. (*Rossumovi Univerzální Roboti - Rossum's Universal Robots*). The writer imagined a world where *artificial people* were built in factories, being so intelligent and so similar to humans that they can be easily mistaken with the latter.

Some years later, about the same time Alan Turing was building the first computers, Isaac Asimov was writing and publishing his famous *Foundation* and *Robot* series, where this concept of a machine that could easily be mistaken for a human and that could perform any task that a human could, and even outsmart the latter, was again used to an even higher extent. Asimov even imagined a series of *laws* a robot should obey at all times so as to protect humans as individuals and humanity as a whole.

Leaping forward to 2019, robots do exist, in various shapes, sizes, and degrees of utility, and slowly begin to become part of our everyday lives. The current definition of a *robot* is, of course, far more complex than what the science fiction authors of the 20^{th} century imagined in their creations. However, the ultimate dream of many people and the ultimate goal of many researchers is to end up developing machines that could perform any task a human can, so as to complement or replace humans from all points of view, much like in the imagination of Karel Čapek or Isaac Asimov.

Robotics is, nonetheless, still a young science. Today's robots can, at best, outperform humans only when dealing with very specific tasks, but they are far from being a pertinent all-around replacement or even complement for humans. This, of course, leaves a lot of room for improvement and many still unresolved challenges for researchers. Ensuring the robots' ability to interact and communicate with humans is one such challenge that has yet to be fully met.

In this thesis, we aim to improve the ability of today's robots to interact naturally and efficiently with humans. This challenge already is a very vast one, but in particular we decided to focus on improving **the ability of robots to detect deception** during the interactions they undergo with humans. In this introductory chapter, we briefly discuss the research context and motivation behind this thesis, the main challenges we have identified, as well as the overview of our research work.

1.1 Research context and motivation

Social robotics focuses primarily on developing the social interaction and communication abilities of a robot. Of course, these are not characteristics that all robots must exhibit, but they are crucial to any robot that is supposed to interact with humans on a regular basis. First and foremost, humanoid robots are expected to possess such interaction capabilities, even though there are even several non-humanoid robots that

are able to communicate with humans to some extent. These abilities make them, on one hand, more appealing to their human users, but also bring them closer to the ultimate goal of robotics: machines that could replace humans in everyday tasks.

Nowadays, social robots have been used in various scenarios in order to improve the quality of life of their users [1] [2]. Besides their use as interactive guides in airports, museums, or stores, as intelligent home assistants or play partners, they have been used with various degrees of success for therapeutic purposes. In particular, **socially assistive robots'** purpose is to improve the quality of life of various populations of users **by means of social interaction** [3].

Several populations of users could and already do benefit from such robots. Generally, the populations of users who would otherwise require long-term and full-time human assistance benefit the most from SAR. Such categories of people are the elderly, people with limited mobility or with chronic disease, people in need of rehabilitation, or with conditions such as autistic spectrum disorders. Social robots can complement or even replace the human assistants who usually take care of these vulnerable users in tasks such as supervising and monitoring them in their home environments or assisting them in various forms of therapy. These tasks involve a constant interaction and communication between the human users and the robots. Since some of the users suffer from medical conditions that impair their communication abilities, the robots' challenge to ensure natural interactions with their interlocutors is even more complex.

1.1.1 Why detect deception in HRI?

As stated previously, it is crucial for socially assistive robots to be able to maintain natural communication and interaction with their human users. This involves being able to understand to the highest possible extent their interlocutors and to express themselves in a way that is easily and clearly understood, to the same extent a human assistant would perform in a similar scenario. However, socially assistive robots face another major challenge that their human counterparts also face: dealing with difficult and uncooperative patients. Depending on their conditions, patients sometimes refuse to undergo parts of their rehabilitation process, take the medication they are prescribed or respect a series of restrictions. Even human assistants have a hard time monitoring such difficult patients and ensuring their well-being, requiring special training to detect and deal with these situations.

In order for a robot to deal with such a challenge, it would also require abilities that are pushing the limits of regular human-robot interaction scenarios. In particular, socially assistive robots would benefit from being able to **detect deception**, so as to be able to maximise their ability to check on their patients' progress. In particular, robots cannot always establish with a high degree of certainty if their human patients accomplished a task required by their therapeutic process, such as exercising, taking certain pills or respecting an interdiction. In these scenarios, they have to be able to ask their users questions concerning the tasks that had to be accomplished and to evaluate both the meaning of their answers (such as confirming or denying the accomplishment of that task), as well as the **honesty** of their answers, at least to the extent a human assistant would be able to.

Moreover, the deception detection process must be as minimally invasive and as fast as possible, ensuring on one hand that the interaction stays as natural as it would be if the robots did not have this ability and, on the other hand, that the human user does not have to be subjected to the use of any sort of cumbersome sensors or devices that would diminish their quality of life or comfort. On the other hand, deception

detection in inter-human interactions does not necessarily imply these restrictions, as it is generally used in situations where establishing the truth is far more important than how invasive the procedure used is, such as criminal investigations. Therefore, the purpose and limitations of deception detection in HRI are significantly different to those of deception detection in inter-human interactions. Consequently, even if our work is strongly based on the previous findings of the research done on the latter topic, our methods are shaped around significantly different objectives.

1.1.2 Challenges of in-the-wild HRI

Being able to detect deception in controlled, laboratory conditions is one thing, but being able to perform the same task in real world conditions is a far more complex goal, especially when using minimally-invasive or noninvasive techniques. For reference, humans unassisted by any deception detection system, are able to detect deception with an accuracy of only 54%, just slightly above that attributed to chance (50%) [4] [5]. In-the-wild HRI implies that, on one hand, the human's movements or behaviour are neither coerced nor restricted in any way and, on the other hand, that the environment conditions, such as lighting, background noise or room temperature, are subject to changes.

Whenever minimally-invasive or noninvasive measurement techniques are involved, there is a trade-off between their measurement accuracy and precision and the variations in the experimental conditions. If humans are allowed to exert their free will and behave as they feel natural and instinctive during HRI, the various behaviours they may exhibit, such as extensive movements or random gestures, the detection or analysis accuracy will be impaired. If, on the other hand, we restrict or coerce the humans into adopting behaviours that would maximise the detection accuracy, then the validity of our measurements would be biased by the fact that we limited the humans' free will, and the physiological manifestations or behaviours we have evaluated may not be their natural reactions when being deceptive or honest.

A similar analysis can be done with respect to the environmental conditions. Having controlled and constant experimental conditions is an important factor in order to ensure that there are no measurement biases induced by any major variations of these parameters from one experiment to the other. On the other hand, if our detection methods and techniques drastically lose their accuracy and precision when being exposed to different conditions, it means they are not usable in real life HRI scenarios, which may take place in various environments with variable conditions. Of course, a certain amount of calibration procedures may be necessary to ensure a given system that ensures the HRI works properly when the environmental conditions change, but the system must be **able** to function properly even when used in situations that are different to those of a laboratory setup.

These challenges and trade-offs have had a major impact on the direction of our research and have shaped our experimental design guidelines. Moreover, they have also influenced our choices of detection methods, algorithms and sensors used throughout this research. Our main objective was that **the methods and results of our work should be used and applied in real HRI scenarios as easily as possible**, even if this meant facing some of the aforementioned challenges associated to in-the-wild HRI. In the following chapters, we will detail and justify our various design, hardware or software choices.

1.2 Thesis overview

The thesis starts with a discussion of the physiological mechanisms of deception, as well as the state of the art deception detection techniques and devices. They are detailed in chapter 2. Then, in chapter 3, we will present the experimental methodology and resources used throughout this work, from the experimental design guidelines we have established to the detailed presentation of the hardware and software resources we have used and developed.

Then, in the following four chapters, we will detail our research work done in order to detect deception in HRI. In chapter 4, we are discussing our first attempts to detect deception, for which we have used only noninvasive measurement devices, chapter 5 details the improvements we managed to make to our initial work by using minimally-invasive devices in conjunction with noninvasive ones, while chapter 6 focuses on the direct comparison between the interaction with a human and the interaction with a robot in a deception detection scenario, using both minimally-invasive and noninvasive techniques. Lastly, our work detailed in chapter 7 is based on a different approach than the one used in the previous chapters, as we attempt to detect deception using only the analysis of the response times to questions.

Ultimately, chapter 8 sums up our results and the list of our publications, while in chapter 9 we will draw the conclusions of this work.

Chapter 2

Physiology of deception

2.1 Introduction

Even though deception detection in HRI is a rather young and unexplored domain, psychologists have attempted and succeeded to some extent to detect deception in inter-human interactions for a few decades already. Therefore, the first source of inspiration for our research came from the previous research conducted in deception detection in inter-human interactions. However, even if we took inspiration for our work from these previous accomplishments, there was no proof that using these previous findings in HRI would yield similar results to those observed in inter-human interactions. On top of that, as we will discover later in this chapter, not all these techniques are actually usable in everyday HRI, the same way they are not usable to detect deception in anything other than a highly controlled environment.

Consequently, we have reviewed the literature in order to establish a preliminary list of deception detection approaches and techniques that were also usable in HRI scenarios. In particular, we were interested in the physiological or behavioural manifestations that have been correlated to deception and that can be measured and evaluated in real time, as the interaction is undergone. Then, we were interested in discovering a variety of measurement devices and methods that could allow us to monitor each of these parameters, in order to find the best compromise between the measurement accuracy or precision and the invasivity of the method.

In this chapter, we start by detailing the two main approaches that have been previously used to detect deception non-verbally in inter-human interactions: the polygraph-based methods and the cognitive load based methods. Then, we present an exhaustive list of manifestations that have been either directly associated to deception, or indirectly, as indicators of an increased cognitive load. W continue by discussing the various devices and techniques used to monitor the previous manifestations, distinguishing between invasive, minimally-invasive, and noninvasive ones. Lastly, we elaborate the differences that have been already observed between the human manifestations when interacting with another human compared to those exhibited during similar HRI scenarios, trying to understand to what extent the nature of the interaction partner could interfere with the ability of detecting deception.

2.2 Approaches to detecting deception

As lie, deception and the concealment of truth have been a part of people's everyday lives since the beginning of humanity, and have always played a crucial role, humans have always been interested in establishing whether their interlocutors were honest or deceitful, at least in key moments. Even though research on this topic has only started in the 20^{th} century, people have attempted to intuitively develop basic

methods for detecting deception many centuries before that. However, according to [5] and [4], studies performed over the past 30 years indicate that the average person is only slightly better at detecting deception than chance, with an accuracy of 54% compared to 50% for chance. Even highly trained individuals, such as police officers, show an accuracy of only 65% in detecting lies and truths on their own, according to [6].

The authors of the book entitled *Detecting deception: current challenges and cognitive approaches* [7] do a very thorough analysis of the past and current methods used for detecting deception, as well as of their limitations. According to Granhag et al. [7], the oldest and most basic lie detection methods are the **verbal lie detection tools**, that establish the veracity of statements based purely on an analysis of their content, completely ignoring any nonverbal, para-verbal, physiological, or behavioural cues. Furthermore, it can distinguish between two main families of methods: the already well-established **polygraph** and all the related approaches, that rely on the evaluation of various nonverbal cues, manifestations and behaviours, and the new, improved model they propose: the **cognitive approach** towards lie detection.

The main objective of past research conducted in lie detection has always been to assist criminal investigators, prosecutors, or law enforcement agencies in their attempts to interrogate suspects or to establish the trustworthiness of people in key positions or situations. Therefore, the emphasis was always put on the **accuracy** of the deception detection techniques and on the ability to establish the truth eventually, and not at all on their usability in everyday interactions or on reducing the time necessary to arrive to a conclusion.

We now detail the two main families of lie detection methods that inspired our work from: the polygraph methods, which have been established a few decades ago and continuously refined since, and the cognitive load based methods, which use a slightly different approach to attempt detection.

2.2.1 Polygraph-based methods

The polygraph is a device monitoring in real time various physiological parameters of a subject, such as their heart rate, respiratory rate, blood pressure or galvanic skin response. According to the authors of [8], who have reviewed the history, methodology and current status of the polygraph, the first such device was created in 1921 by John A. Larson. The device monitored initially only the heart rate, respiratory rate and blood pressure, and was later improved by Leonarde Keeler, who made it portable, added the measurement of the GSR and also patented the device.

In parallel, various researchers and people working in law enforcement agencies have improved both the devices and the techniques and procedures used to interrogate the subjects who were connected to the device. One of the main figures of the future development of the polygraph [9] was John E. Reid, who developed the Comparative Question Test (CQT) examination procedure [10] and was also interested in evaluating the reliability of the device and the methodology [11], as it started to be used in criminal investigations and by police forces.

The reliability of the polygraph device and of the associated interrogation methods has been, however, strongly contested, in spite of a number of studies confirming a very good detection reliability. In the initial study done by Horvath and Reid in 1971 [11], the percentages of correct judgements made by examiners who evaluated polygraph records varied between 85% and 97.5% in the case of experienced examiners, and were still over 70% in the case of inexperienced examiners with less than 6 months experience. Moreover, according to a later study made by Andrea

Gagiolli [12], modern polygraphs can achieve a detection accuracy of 81% to 91%. Nonetheless, Iacono and Lykken [13] have concluded that polygraph lie detection has no theoretically sound basis supporting it, that it can be easily cheated using the appropriate countermeasures and that it should not be admitted as evidence into courts of law. They have based their conclusions on two surveys filled by scientists that were members of the Society for Psychophysiological Research or Fellows of the American Psychological Association, so their results can also be considered subjective.

Besides the polygraph device in itself and the debates regarding its efficiency, researchers have attempted to establish other techniques that would allow them to detect deception using the same basic principle: the correlation between the evolution of one or several physiological manifestations and the veracity of a person's statements or answers. These alternative approaches are still considered polygraph methods or polygraph-based methods, as they work in a similar fashion. Some of these approaches are even more invasive than the polygraph device, involving for example the monitoring of the brain activity using Functional Magnetic Resonance Imaging [14], while others are less invasive and can even be performed remotely. These latter, less invasive, attempts to detect deception have been of particular interest for us, some of them using thermal facial screening [15] [16], the analysis of blinking patterns [17], the analysis of behavioural cues [18] or of pupillary dilation [19]. Other polygraph based lie detection methods simply monitor one or several of the physiological parameters monitored by the polygraph device, using more common and less invasive measurement devices.

To sum up, the polygraph device is considered today the state of the art deception detection technique, and has been established as such for a few decades already, in spite of a certain amount of debate surrounding the theoretical validity of the device and of the technique. However, since the device is highly invasive and requires controlled interrogation techniques and environments, it is virtually unusable in everyday HRI (or in everyday inter-human interactions, for that matter). However, the main principle behind the polygraph, which is the correlation between the evolution of physiological manifestations and the truth value of one's statements or answers, is of particular interest for our research, and same is to be said about all lie detection techniques that involve a polygraph approach. In section 2.3, we present a more exhaustive list of physiological manifestations that have been correlated to deception, mainly in inter-human interactions, but also in HRI.

2.2.2 Cognitive load based methods

As previously mentioned, the polygraph and polygraph based lie detection techniques are highly contested by some scientists. On top of that, as concluded by Bond and DePaulo in [5], the average person's ability to classify truths and lies accurately is of only 54%, while some individuals, either highly trained or with a certain *natural ability* to detect lies, being able to achieve significantly higher classification rates. However, since neither the polygraph nor the innate or educated human lie detection skills are scientifically sound or based on a theoretically sound basis, recent research in lie detection has attempted to find a new, more rigorous lie detection system.

An alternative concept has been shaped in 1981 by Zuckerman et al. [20], who concluded that the action of lying increases the cognitive load, or the amount of thinking required by individuals, with respect to telling the truth, and that this increase of the cognitive load comes with a series of observable signs. The list of these

signs will be discussed later, in section 2.3. The main principle of this concept is that instead of trying to evaluate physiological or behavioural manifestations that are supposed to be directly correlated to deception, the manifestations that indicate an increased cognitive load should be looked at, and that these manifestations can be observed more easily and more accurately.

This model was later improved by Vrij et al. [21] [22], who suggested that artificially increasing the subject's cognitive load throughout the interrogation would amplify even more the differences that can be observed between the cognitive loads of liars and those of truth tellers. This moved the emphasis on the development of various interrogation techniques that achieve the goal of maximising the differences between the cognitive loads exhibited when lying compared to telling the truth, instead of focusing on the direct analysis of various physiological manifestations. Several such techniques have been proposed: Vrij et al. proposed in [22] the use of unanticipated questions, the *Devil's-advocate approach* or the strategic use of evidence, then later proposed in [23] an interrogation procedure where events are recalled in a reverse order so as to increase the cognitive load. Many other alternative cognitive load inducing techniques have been developed since, and they have been reviewed and compared with the traditional polygraph approach in [24], validating them as alternatives to the polygraph and the Comparative Question Test, and even as ways of averting the mistakes and pitfalls of the polygraph based approaches.

From our perspective, the main issue with the cognitive load approaches is, again, that they are not appropriate for everyday HRI interactions, especially the latter models that rely on *imposing* a cognitive load by using very specific interrogation procedures. Much like the polygraph-based approaches, their main purpose is establishing the truth with the highest possible degree of accuracy, and not being able to use them in regular inter-human or HRI interactions. Therefore, as efficient and accurate as some of these techniques may be, they cannot be used outside some very specific interrogation scenarios, rendering them unsuitable for our research.

However, as it will be discussed in section 2.3 and section 2.4, some of the physiological manifestations that are associated with an increased cognitive load can be easily monitored during in-the-wild HRI scenarios, using noninvasive or minimally-invasive methods. This would allow us to evaluate the cognitive load of subjects, regardless of the interaction scenario, and then to attempt to detect deception based on the cues indicating an increased cognitive load.

2.3 Manifestations associated with deception

The main deception detection methods we have described in the previous section rely on the evaluation of several human manifestations. In particular, polygraph devices and polygraph based methods rely mainly on the evaluation of raw **physiological** manifestations as well as **behavioural** ones. On the other hand, cognitive load approaches focus both on verbal and para-verbal manifestations, as well as physiological or behavioural ones.

In this section, we present a list of manifestations that have been systematically monitored in order to detect deception, as studied previously in inter-human interaction scenarios. As most of our research focuses on polygraph based approaches, our analysis of manifestations commonly used in these approaches will also be more thorough.

2.3.1 Physiological manifestations

The polygraph device monitors a series of physiological manifestations whose evolution is at least partly influenced by the autonomic nervous system (ANS), that automatically regulates internal body processes, such as the blood pressure, heart rate, respiratory rate, body temperature, sweating or arousal. The ANS is composed of two subsystems: the sympathetic nervous system and the parasympathetic nervous system, the former being in charge of preparing the human body for a variety of emergency situations, or the so-called *fight or flight* response [25]. Many different situations trigger the *fight or flight* response, including deception, since it usually involves hiding a truth whose reveal would have negative consequences for the person hiding the truth.

After reviewing the literature, we have compiled a list of physiological manifestations whose evolution and variations have been correlated to deception. Some of these manifestations are evaluated by polygraph devices, while others and their correlation with deception has been studied independently of the development of the polygraph device. Last, but not least, some of them have also been correlated to an increased cognitive load. The list of physiological parameters we were interested in is the following:

- heart rate
- blood pressure
- respiratory activity
- galvanic skin response
- skin temperature
- pupillary dilation
- brain activity

Heart rate and blood pressure

The heart rate and blood pressure were among the parameters monitored by the first polygraph devices and they were monitored ever since. In their book entitled *Truth and deception: the polygraph ("lie-detector") technique* [9], Reid and Inbau have defined the methodology used to evaluate the veracity of a person's answers using the polygraph device. Their methodology indicates that a person's heart rate will increase more when giving a deceitful answer to a question than when giving an honest answer. Later studies [26] [27] [28] have also validated the model proposed by Reid and Inbau, stating that the detection rates obtained using heart rate measurements, although variable, are greater than chance (therefore greater than 50% accuracy rate).

Blood pressure was also considered as a potential indicator of deception as early as 1917, when Marston [29] concluded that sharp increases of the subjects' blood pressure are correlated to lies during story telling, but also that increases in blood pressure are mainly induced by fear and anger. Reid also acknowledged the correlations between the blood pressure variations and deception [30], but also concluded that those variations can be artificially induced by the interrogated subjects by contracting their muscles and constricting their blood vessels. Reid therefore acknowledged the existence of countermeasures that *trained* deceivers can employ in order

to cheat the polygraph, but also proposed an instrument that would detect the voluntary muscular movements that participants may use in order to falsify their blood pressure measurements.

Respiratory activity

Respiratory rate and amplitude are both indirectly controlled by the sympathetic component of the autonomic nervous system, and as triggering the *fight and flight* response dilates the bronchioles and increases the lungs' ability to exchange oxygen and carbon dioxide [25]. However, respiration is also controlled by the central nervous system, which means that it can be easily regulated consciously by individuals. Nonetheless, since the first polygraph devices were developed, they also monitored respiratory rates and amplitudes. In particular, according to the methodology of Reid and Inbau [9], the subjects' breathing amplitude decreases when they are dishonest compared to when they are honest, and the respiration rate also increases when participants are dishonest compared to when they are honest.

Studies evaluating the accuracy of respiratory activity as a lie detection technique have shown rather contradictory results. If Cutrow et al. have observed an even higher detection accuracy compared to the one obtained when monitoring the heart activity [31], Podlesny and Raskin have not found any correlation between the respiration amplitude or the respiration cycle time and the truth value of the subjects' answers [26]. Therefore, despite being one of the parameters monitored by polygraph devices, the respiratory activity is not necessarily a reliable measure when used on its own in order to detect deception, and is also one of the easiest manifestations to control consciously by subjects who employ countermeasures during the interrogation process.

Galvanic skin response

Another physiological manifestation triggered by the *fight or flight* response of the sympathetic system is stimulating the sweat glands to produce perspiration [25]. Since perspiration is rich in minerals, including salt, and therefore a good electrical conductor, while human skin is a rather poor electrical conductor in comparison, even the slightest amounts of perspiration have a major impact on the skin resistance or conductance. Therefore, Skin Conductance (SC), Skin Resistance (SR) or, more generally, Galvanic Skin Response (GSR), which are different measures describing the variation of the skin's propriety to conduct electrical current due to perspiration, have been considered a reliable cue to deception, also being monitored by polygraph devices. As Reid and Inbau [9] have established, the subjects' skin conductance increases more (or, consequently, their skin resistance decreases more) after giving a dishonest answer compared to giving an honest answer to a question.

Several other studies have tested this model, attempting to predict the value of the participants' answers based on the variations in their SC or SR. The authors of [28] have observed that, when using the Guilty Knowledge Test as an examination method, the SR measures have been the most effective in correctly distinguishing between truths and lies, compared to the other physiological parameters they have evaluated (cardiovascular activity and pupillary measures). Cutrow et al. [31] have done a similar comparative study, also obtaining a prediction accuracy based on the SR superior to that of the prediction accuracy based on the HR. Lastly, Podlesny and Raskin [26] have studied several measures of the SC (its amplitude, its rise time, its

recovery half-time and its recovery half-time width), in order to analyse the amplitude variations of the SC as well as the speed with which it varies. They have also confirmed that the amplitude of the SC variation is higher for guilty subjects compared to innocent ones. The guilty subjects' recovery half-times were also higher compared to innocent ones.

Skin temperature

One of the more subtle effects of triggering the *fight or flight* response is the increased blood flow in the ocular areas. According to Pavlidis and Levine [15] [32], one of the effects of the *fight or flight* response is to facilitate rapid eye movement by increasing the blood flow irrigating the ocular muscles. An increased blood flow that can be localised in the ocular area would, therefore, increase the temperature of the skin around the eyes (the periorbital region). Similar research was undergone by Pollina et al. [33], Sumriddetchkajorn and Somboonkaew [16] and Rajoub and Zwiggelaar [34], all observing similar increases of the periorbital temperature when participants were deceitful.

Moreover, the authors of [16] and [34] have attempted to develop autonomous deception detection systems based on the measurements of the periorbital area temperatures, using Machine Learning algorithms, each obtaining an overall detection accuracy of 84% and 62%, respectively. These values are in line with those obtained using traditional polygraph devices.

Pupillary dilation

Pupillary dilation, even though included in this list of physiological parameters that have been correlated to deception using the traditional, polygraph-based approach, has also been correlated to an increase in the cognitive load [7]. Pupillary dilation, which allows for more light to enter the eye and therefore to improve the eye sensibility to light, is undoubtedly controlled by the sympathetic nervous system when the *fight or flight* response is triggered [25]. However, Kahneman established in 1973 [35] that pupillary size is also an indicator of an increased cognitive load, his study being one of the basis for future research on the correlation between the dilation of pupils and deception.

Whether it is directly correlated to deception or indirectly, via the increased cognitive load, several later studies have shown the correlation between pupil dilation and deception, such as those undergone by Heilveil [36], Janisse and Bradley [37] [38], Lubow and Fein [39] and Seymour et al. [40]. Moreover, pupillary dilation was one of the few physiological manifestations that have been also studied when attempting to detect deception in HRI scenarios. In particular, Aroyo, Gonzalez-Billandon et al. have attempted to detect deception in HRI [41] [42] using, among other manifestations and measures, the pupillary dilation. They have observed a similar correlation to that previously established in inter-human interaction, with deceptive people exhibiting higher pupillary dilation than truth tellers in an HRI scenario.

Brain activity

One of the hardest to evaluate physiological parameters that has been correlated to deception is the brain activity, using functional Magnetic Resonance Imaging (fMRI). Several studies [14] [43] [44] have proven that different areas of the brain are being

activated when telling the truth and when lying, and their activity can be monitored using fMRI. Moreover, the authors of [45] have successfully trained an autonomous lie detection system, based on fMRI and Support Vector Machines (SVM), obtaining a detection accuracy of over 88%, significantly better than most polygraph devices. However, the procedure and the devices necessary for implementing this lie detection technique are highly invasive, as the patient has to be monitored using an MRI scanner. Therefore, as promising as this lie detection technique is, its use in HRI is completely out of the question.

Summary

The various physiological manifestations we have discussed previously have been successfully used as indicators of deception in past research, mostly in inter-human interactions, but also in HRI in the case of pupillary dilation. Their accuracy when used independently varies quite significantly according to previous studies, much as it is the case for the polygraph device, but from this point of view we can still consider them viable candidates in order to detect deception in HRI. However, measuring some of these parameters in in-the-wild HRI scenarios is highly limited by the necessary measurement devices. The obvious example is that of the brain activity, which requires the use of MRI devices, making it completely unsuitable for our research. The use of the others, as we will discuss in section 2.4, will often involve a compromise between the measurement accuracy and the invasivity of the measurement devices.

2.3.2 Behavioural manifestations

When discussing the correlations between physiological manifestations and deception, the main assumption was that the variations of each parameter were triggered by the sympathetic nervous system and, with the exception of the respiratory activity, they are very hard or even impossible to regulate using the central nervous system. Behavioural manifestations, on the other hand, involve mainly the use of skeletal muscles, which are significantly less under the control of the autonomic nervous system, and they are also easier to regulate consciously by using the central nervous system. This makes most behavioural manifestations easy to control consciously by subjects who want to employ countermeasures against deception detection.

However, a person's behavioural manifestations can be evaluated remotely, either by trained individuals making manual annotations of the observed behaviours or by using video cameras and tracking algorithms that evaluate the behaviours autonomously. This is what makes the evaluation of behavioural manifestations more appropriate for detecting deceptions in *in-the-wild* HRI scenarios. After reviewing the literature, we have compiled the following list of behavioural manifestations that have been correlated to deception:

- body posture
- head and gaze orientation
- response times to questions
- blinking

Some of these manifestations have been directly correlated to deception, while others have been correlated to an increased cognitive load, making them more suitable for cognitive load based deception detection techniques. In this subsection, we

will detail how and to what extent each of these manifestations have been considered reliable cues of deception in inter-human interactions as well as HRI.

Body posture

Researchers have shown that liars deliberately self-control their behaviours, as well as their thoughts and feelings when lying, to a greater extent than truth tellers [46] [20]. This self-regulatory effort has an effect, on one hand, on their mental resources, increasing their cognitive load, but also on details of their behaviour such as their body posture or the movements made with their hands or legs. According to DePaulo et al. [46], the general social perception is that liars could not stand still and that they would fidget, shift their posture a lot or shake their legs. Therefore, as a counter-measure employed in order to avoid exhibiting such behaviours, liars tend to self-regulate their behaviour more than a truth teller would do in a similar context.

There are several indicators of a self-regulatory behaviour, starting with the body posture. Several studies undergone in inter-human interactions [47] [48] [49] have proven that participants have a more controlled body posture when lying than when telling the truth. The more controlled body posture of the participants, who were sitting down during the experiments, meant less head and hand movements, as well as a more centred and upright position, with their head placed equidistantly with respect to their hands. The subjects' body posture was evaluated by means of automated video analysis, with algorithms tracking the head and hand positions relative to the frame and then computed several metrics based on their positions and speeds.

Head and gaze orientation

Gaze is an important non-verbal component of inter-human interactions, especially maintaining or avoiding direct eye contact with one's interlocutor. If gaze is defined as the direction towards which a person's eyes are pointing, it can be decomposed into two components: firstly, the person's head position and orientation, and secondly, the eyeball orientation with respect to their head. Therefore, maintaining eye contact for sustained amounts of time involves orientating their heads towards their interlocutors, while gaze aversion also involves a head posture that is not oriented towards the interlocutor.

If we apply the theory we have discussed previously, according to which liars tend to self-regulate their behaviour more than truth tellers, they would be expected to also maintain a more natural and constant gaze orientation throughout the interaction compared to truth tellers. However, according to Loy et al. [50], several studies analysing the correlation between maintaining or averting direct gaze and honesty have contradicted this hypothesis or have observed inconsistent behaviours. In two studies undergone by Mann and Vrij [6] [51], trained police officers who were asked to evaluate the veracity of suspects' statements have associated gaze aversion with a deceptive behaviour. Moreover, Vrij and Mann [51] have observed that the same subject has shown more gaze aversion when lying compared to truth telling during one of the interviews and exhibited the opposite behaviour in another one. Lastly, two meta-analytic studies, the first by DePaulo et al. [46] and the second by Sporer and Schwandt [52], conclude that there is no correlation between deception and gaze aversion or maintaining eye contact.

Therefore, according to Loy et al. [50], eye contact and gaze aversion may be also attributed to situational variations, such as different questioning styles, the complexity of the lies involved, or even to what extent the participants were instructed to *lie*

on cue instead of exerting their free will. In spite of these inconclusive or contradictory results obtained by the various studies undergone in inter-human interactions, we have decided to consider the head position and orientation as a parameter that could potentially be correlated to deception in HRI, as it is considered both an indicator of the body posture, as previously discussed, and of the gaze.

Response times to questions

As previously mentioned, in section 2.2, one of the main methods to detect deception is the cognitive load based approach. The main principle behind this approach is that making deceptive statements or giving deceitful answers increases the cognitive load of the subject compared to making honest statements or giving honest answers. One of the indicators of an increased cognitive load is the time required to answer the interlocutor's questions, an increased cognitive load being associated to higher response times. Previous studies have confirmed this theory both in inter-human interactions [20] [21] [53] and in HRI [41] [42].

Measuring the response times to questions is a very easy and straightforward procedure, requiring a minimum of devices, such as a microphone and an audio recording system. Traditionally, it is done by manually measuring and annotating the response times, but that can be also easily automatised so that response times are extracted in real time. Its very low degree of invasivity and its ease of implementation makes it a viable candidate method to detect deception in HRI scenarios.

Blinking

Blinking is another behavioural manifestation that has been correlated to the cognitive load. Firstly, a series of studies have proved that the blink rate decreases during the time when the subject's cognitive load increases [54] [55] [56], while others have also proved that during the cognitive break occurring after a period of increased cognitive load, the subjects' blink frequencies increase significantly [57] [58]. Applying this correlation to the previous assumption that lying increases the cognitive load of subjects compared to telling the truth has lead to the theory that subjects would blink less when lying compared to telling the truth, and also that their blink frequency would increase more after telling a lie than after telling the truth.

Some of the researchers who studied this correlation between blink rates and deception has validated this theory, while others proved the contrary. One one hand, Leal and Vrij [17] have observed that liars have indeed reduced their blink rates when lying and significantly increased them afterwards, with respect to the baseline, while truth tellers exhibited a somewhat opposite pattern, with a slight increase of the blink rate while telling the truth. On the other hand, DePaulo et al. [46] have observed an opposite behaviour, with liars blinking more than truth tellers. The correlation between the blinking frequencies and deception was also studied in HRI by Aroyo, Gonzalez-Billandon et al. [41] [42], but they haven't found any statistically significant correlation between the participants' blink rates and the veracity of their answers. Therefore, in spite of the theories indicating that blinking rates and patterns should be a valid indicator of deception, previous studies have contradictory or inconclusive results.

Summary

As we can conclude so far, behavioural manifestations are not the most reliable indicators of deception, and they are significantly less reliable than physiological ones.

This can be explained, on one hand, by the far greater complexity of behavioural manifestations compared to the physiological ones, and on the other hand by the fact that the central nervous system has a significant degree control over our behaviour, making it much easier to regulate than our raw physiological reactions. Lastly, a study by Vrij et al. [59] concludes that one of the reasons for which many studies focusing on the correlation between behavioural cues and deception fail to arrive to coherent conclusions is that they are examining the *wrong cues*.

Nonetheless, as mentioned in the introductory part of this subsection, the study of physiological manifestations is of a particular interest for us, mainly due to the fact that we can evaluate these manifestations remotely and noninvasively. If the physiological manifestations we have discussed in the previous subsection usually require the use of devices that involve a certain amount of contact and interaction with the human body, behavioural ones can be easily evaluated using either external observers (such as trained individuals who monitor various cues) or by means of computer vision and tracking algorithms. And since all the behavioural manifestations we have discussed here (body posture, head and gaze orientation, response time to questions as well as blinking patterns) can be evaluated automatically by computer systems, sometimes even more accurately than trained human annotators, they will be undoubtedly of major interest for our research.

The following table summarises the manifestations associated to deception, as they have been studied previously in literature:

TABLE 2.1: A review of manifestations associated to deception

Parameter	**Values associated to deception**
heart rate	increases to higher values
blood pressure	sharply increases
respiratory activity	rate increases, amplitude decreases
skin conductance	increases to higher values
skin temperature	increased skin temperature in the periorbital region
pupillary dilation	higher dilation
body posture	over-controlled posture, less movements
head and gaze orientation	gaze aversion (though debated by some)
response time to questions	longer time necessary to answer
blinking rate	lower rate when lying, significantly higher rate afterwards

2.4 Measurement devices and techniques

In the previous section, we have discussed the correlation between various physiological and behavioural manifestations and deception, as they were studied previously in literature. Besides understanding what are the cues that indicate that a person is trustworthy or deceitful, it is also important to understand **by what means we can monitor** those cues and to what extent the person whose honesty we are trying to establish can interact, function and behave normally.

If behavioural manifestations are generally monitored from an external point of view, either by trained humans who evaluate and annotate manually various behavioural cues, or by computer systems using image and audio processing, physiological manifestations require devices that tend to be significantly more invasive than that. Many of the devices traditionally used to monitor human physiological manifestations are the ones used in hospitals, or using technologies similar to those

used in medical devices. The design of such devices hardly ever takes into account to what extent the patient (or subject) can perform everyday tasks while being monitored.

Therefore, before attempting to implement any form of deception detection in HRI, we must also understand how we can monitor each of the physiological parameters of interest to us. In particular, we focused on finding out what measurement devices and techniques allow us to make measurements that are accurate and fast enough for the deception detection process, but with the lowest possible degree of invasivity. We have distinguished between three main types of devices: invasive, minimally-invasive and noninvasive ones. In the next part of this section, we will detail the various state of the art devices and techniques that fall under each category.

2.4.1 Invasive measurement devices

Traditionally, most of the physiological manifestations we have previously discussed, such as the heart rate, blood pressure, respiratory activity or GSR, are measured by invasive means. These invasive measurement devices imply the use of a certain amount of sensors placed on the subject's body, connected using wires to the main device analysing the data. State of the art physiological measurement devices focus on achieving the highest possible degree of measurement accuracy and precision, as they are generally used for medical applications.

In particular, the polygraph device, even in its latest iterations, also uses invasive sensors to monitor the subject's physiological parameters. Moreover, the use of polygraph devices involves seating down and constraining the subject, and instructing them to stay as still as possible. According to the analysis done by Synnott et al. [8], most polygraphs use the following types of sensors:

- the cardiovascular activity (heart rate and blood pressure) is monitored using a sphygmomanometer arm cuff, that sometimes also comes as a wrist cuff or finger cuff device, or photoelectric plethysmographs, that are placed either on ear lobes or fingertips
- the respiratory activity is measured using pneumatic rubber bellows fastened around the thorax and/or abdomen, or piezoelectric transducers
- skin conductance is monitored using two electrodes placed on the subject's palm or fingers
- extra sensors are used to monitor the so-called *activity* of the subject during the interrogation, such as their muscle movements

To these, we can also add the most invasive device of them all, the fMRI scanner mentioned in the previous section. At this point, it becomes obvious that invasive measurement devices are unsuitable for detecting deception in HRI scenarios, in spite of offering the highest possible degree of measurement accuracy and precision. Their use generally requires the subject to be sitting or lying down, and also not to move outside a given range, limited by the cable connection between the sensors and the measurement device.

However, given that some of the fluctuations of the physiological parameters that are associated to deception are subtle and short term, and that most of the research on this topic was done using such invasive sensors and devices, it is important that we use measurement devices that can offer a comparable measurement

precision, accuracy, resolution and response time. In particular, except for the blood pressure measurements, these devices offer measurements in real time, with no significant delay with respect to the detected event (individual heartbeats, amount of expansion of the lungs, changes in skin conductance, or even brain activity in case of fMRI) and with very low measurement errors.

For this reason, when discussing the various minimally-invasive and noninvasive measurement devices that can perform similar tasks, we will compare their performances to the benchmark set by the invasive measurement devices that can be used instead.

2.4.2 Noninvasive measurement devices

At the opposite end of the invasivity scale, we find the noninvasive measurement devices and techniques. The main advantage they have, by definition, is that they can measure physiological parameters remotely without being in physical contact with the subject, relying mainly on video cameras (RGB, RGB-D or thermal) and image processing techniques. We will now discuss the list of parameters and their corresponding tracking methods based on the device used to monitor them.

RGB / RGB-D cameras

Regular RGB video cameras, as well as RGB-D ones, are the main devices used for automatically evaluating a subject's behavioural manifestations. Using video streams, sometimes complemented by the depth information provided by RGB-D cameras (Microsoft Kinect, Asus Xtion or others), tracking libraries such as the CLM [60], CLM-Z [61] or DLib [62] allow for the tracking of facial features, head position and orientation, blinking or body posture. On top of that, the microphones integrated in the cameras allow for easy voice analysis, including the automatic detection of response times. These cameras do not have to be placed very close to the subject, meaning that they can be embedded into the robotic platforms and, with appropriate fields of view and sensor resolutions, provide useful video streams from regular interaction distances between the robot and its subjects.

Furthermore, video cameras and computer vision methods can be used in order to track the gaze orientation, provided that they are placed closely enough to the subject's eyes. Dedicated eye tracking devices that do **not** have to be worn by the subjects, thus being completely noninvasive, are readily available, such as the SR Research Eyelink-1000 binocular-arm used by Wang et al. [63]. However, estimations of both the eye gaze [64] [65] can be obtained using regular RGB video cameras and various tracking algorithms. Nonetheless, both dedicated and custom developed systems need to be placed very close to the subject in order to ensure accurate and precise measurements, which imposes a number of restrictions on the nature of the HRI scenarios, restrictions of both interaction distance and the allowed amount of movement of the subject being monitored.

Lastly, RGB cameras can be used to monitor a subject's heart rate using the same principle of photoplethysmography (PPG) used in fingertip heart rate sensors. As the blood flows through the body and irrigates the skin at each heart beat, the skin's colour varies slightly due to the blood's haemoglobin content. This colour variation is imperceptible to the human eye, but it can be detected using image processing techniques. More specifically, an algorithm identifies and tracks a fixed region of skin, such as the forehead, and at each frame computes its average green colour coordinate of the RGB coordinate system. Then, the heart rate is extracted either by

means of FFT analysis or by peak detection. In case of FFT analysis, the generated data series in the time domain is converted to a frequency spectrum, which is then restricted to the frequency bands that correspond to valid human heart rates (typically 1...3 Hz). The most significant frequency peak thus obtained corresponds to the median heart rate during the measurement window.

This method was already used in literature in order to monitor human heart rate remotely, by Verkruysse et al. [66] and by Poh et al. [67] [68]. Other alternative PPG methods have also been used previously, involving either peak detection [69] or other signal processing techniques that did not involve frequency domain analysis [70]. Even though it is reliable in ideal conditions, such as the subjects not moving or not exhibiting sudden variations in their heart rates, this technique has its drawbacks. FFT analysis based measurements require the use of 30 seconds analysis windows, or even longer, meaning that the measurements cannot be made in real time. Moreover, only the median heart rate of the entire analysis window can be extracted, which means that any heart rate variations during the analysis interval will not be detected. Lastly, methods relying on time domain analysis, that filter the PPG signal and detect peaks in real time, are very sensitive to noise sources, such as variations of the lighting conditions or movements of the subject's head.

Thermal cameras

Firstly, the **skin temperature** and, more importantly, the skin temperature of very specific facial areas, is significantly easier and more accurately monitored using thermal cameras than any other more invasive measurement device. A thermal camera works very similarly to a regular, RGB video camera, but instead of measuring the colour and brightness of each point in its field of view, it measures its temperature. Thermal cameras have been already used when researching the relationship between skin temperature and deception [15] [16] [32] [33] [34], but also in order to detect other psychological states, such as anxiety [71], stress [72], mood while playing a game [73] or when attempting to measure the cognitive performance [74].

The main issue with the use of thermal cameras for measuring the temperature of specific areas of skin is the lack of readily-available facial tracking libraries for thermal imaging. Therefore, in order to track the temperature of a specific region of skin, there are four major alternatives, each with its own strengths and weaknesses:

- manually annotating the coordinates of the regions of interest in the thermal frames, frame by frame: very time consuming
- requiring the subjects to stay completely still and tracking the temperatures of the same regions in every frame: fast, easy to implement, but not appropriate for natural, in-the-wild interactions
- using an RGB camera in conjunction with the thermal one, tracking the facial features in the RGB frame using libraries such as CLM [60] or DLib [62], and then converting their coordinates to the thermal frame: a reasonable compromise between ease of implementation, versatility and measurement accuracy
- training a facial feature tracker based on thermal images: the most accurate, but also the most time consuming solution

Going one step further, and using the same principle of tracking the temperature of a given region of the facial skin, a subject's heart rate and respiratory rate can be

monitored remotely and noninvasively. More specifically, as blood flows periodically through the skin, at each heartbeat, the skin's temperature fluctuates periodically. Similarly, as a person breathes in and out, either on their nose or their mouth, the temperature of the region situated in front of those regions fluctuates periodically. If a tracker monitors at each frame the temperature of each region of interest (a region of facial skin, or the area in front of the nostrils or the mouth), the frequency of these periodical variations can be extracted either using FFT analysis techniques or in real time by detecting local peaks. This has already been attempted in literature, by Pavlidis et al. [75] (heart and respiratory rate), Garbey et al. [76] (heart rate), Chauvin et al. [77] (respiratory rate) and Rumiński [78] (respiratory rate).

As promising as these techniques sound, they have some major limitations when it comes to the measurement accuracy, resolution and speed. First of all, the techniques relying on FFT processing [75] [76] [77] require the analysis of windows of samples of at least 30 seconds, and even as long as 100 seconds, which will be significant measurement delays. Moreover, the computed heart rate and respiratory rate values will be the median ones throughout the interval covered by the sample window, and any variations occurring during that window will not be perceived, as individual heart beats or respiratory events cannot be detected. Lastly, the alternative represented by peak detection methods, such as the one presented by Rumiński [78], is much more sensitive to noise than the FFT-based methods. Such sources of noise are represented by movements (and especially the rotation) of the head or speaking, which will easily occur during HRI scenarios.

Summary

To conclude, camera-based noninvasive measurement techniques offer a pertinent alternative to the more traditional devices used to monitor the human physiological activity. Moreover, since they rely on RGB, RGB-D or thermal video cameras, which are generally part of modern robotic systems, they seem like a reasonable choice for our research work in HRI. Nonetheless, they are subject to sources of noise, they do not offer the same measurement precision or accuracy outside ideal measurement conditions, and they are also mainly unable to estimate the measured values in real time, as FFT analysis requires measurement windows of at least 30 seconds.

2.4.3 Minimally-invasive measurement devices

Minimally-invasive measurement devices attempt to offer measurement accuracy and precision that are close to those the invasive devices, while still allowing the user to behave and interact naturally and to perform common tasks unhindered. The main difference compared to noninvasive devices is that they still require a number of sensors or a part of the device to be in contact with the user, who will carry around the device or a part of it in a convenient and minimally intrusive way. When it comes to the comparison of the use cases allowed by each type of device, both minimally-invasive and noninvasive devices clearly allow for a significantly higher number of use cases compared to the invasive devices. Directly comparing the minimally-invasive and noninvasive devices directly isn't as straightforward though, as minimally-invasive devices sometimes allow the user to perform even more tasks than noninvasive ones, at the cost of wearing the device.

With the advent of computing technology and especially of the Internet of Things (IoT), various forms of wearable devices have become not only available, but also quite ubiquitous in our lives. The most prominent example is the smartwatch, even

the most basic ones now incorporating PPG heart rate sensors and accelerometers. Smart glasses with eye tracking capabilities are also being available on the market today, albeit not being nearly as popular as smartwatches. The advent of wearable devices has also shown that humans become more and more open to the idea of wearing devices that monitor their physiological activity or their behaviour, as long as they improve their quality of life in a reasonable way.

As PPG sensors are already integrated in smartwatches or smartbands, heart rate is one of the main physiological parameters that can be evaluated by minimally-invasive means. These sensors measure the heart rate by detecting the flow of blood through the skin by optical means, using an LED and a photosensitive sensor, usually in real time and with a good accuracy and precision. As the sensor is in contact with the person's wrist, there are very few sources of noise or measurement error. Wearable PPG sensors have been used in research [79] even before smartwatches became overly popular, and have proven to be reasonable alternatives to more invasive measurement devices. Other, more traditional ways of monitoring the heart rate or respiratory rate with minimally invasive devices involve the use of chest bands, which measure either the electric potential of the skin in order to detect individual heart beats [80] [81] or the expansion of the rib cage with strain or stretch sensors [81]. Similarly, GSR can be also measured using minimally-invasive wearable sensors, such as the models manufactured by Eisco, Mindfield or Shimmer, and such systems have already been used in research to analyse physiological manifestations [82] [83]. Therefore, all the major parameters analysed by the polygraph machine by invasive means can also be analysed by minimally-invasive means.

Lastly, minimally-invasive devices in the form of smart glasses are the most appropriate method to evaluate an user's pupillary dilation, eye gaze and blink rate. Such dedicated smart glasses are offered by several manufacturers, such as Arrington, iMotions, Pupil Labs, Tobii Pro or SR Research, and have been already used in research studying the correlation between pupillary dilation or gaze and deception [40] [41] [42]. Since monitoring the ocular activity requires placing a video camera as closely as possible to the human eyes and limiting its movement relative to the tracked eyes, smart glasses with integrated video cameras are by far the most reliable and accurate solution to track pupil dilation, gaze or blinking.

Summary

As we can see, minimally-invasive measurement devices seem to represent the *jack of all trades*, offering the possibility to measure several physiological parameters in real time and with a good reliability, and in some cases the best reliability, while still allowing the wearer of these devices to behave and interact with minimum restrictions. Nonetheless, their use as complementary tools in HRI and especially in deception detection scenarios has one major flaw: the user can, at any moment, choose not to wear the devices, thus rendering the robot unable to monitor the physiological activity those sensors are supposed to evaluate. Therefore, for the deception detection process to work as intended, the human user must "cooperate" with the robot by wearing all the required sensors or devices, or to be coerced into wearing the sensors at all times without being told they are used by the robot to detect when they are deceitful. In both cases, the user has an easy option to render the robot unable or less able to detect whether they are deceptive or not if they choose to, which may be a major pitfall in some interaction scenarios.

2.5 Applying theories of inter-human interactions to HRI

After studying and summing up the conclusions of the past research done in deception detection in inter-human interactions, one question was also implicitly raised: **can we assume these results are also valid in HRI scenarios**? In order for it to be the case, we need solid proof that humans perceive social robots similarly to humans during interactions and also that their physiological and behavioural manifestations are not influenced by the nature of their interaction partner. Very few studies trying to directly compare inter-human and human-robot interactions have been undergone so far, especially since social robotics is a quite young science and since social robots are still far from being able to interact and communicate as naturally as a human. However, the studies undergone so far tend to prove the contrary.

For example, Poore et al. [84] have compared inter-human interactions with interactions undergone with virtual agents, measuring various physiological parameters during the interactions, and concluded that the participants heart rates were higher when interacting with virtual agents instead of human interlocutors. Then, Willemse et al. [85] studied the human response to robot-initiated social touch, as past studies had shown that stress level decreases when humans are being briefly touched by other humans on their arm during interactions. The authors of the study have concluded that, when humans were touched by robots instead of humans, no such decrease was observed.

Lastly, research undergone in our laboratory by Agrigoroaie, Tapus and Cruz Maya has investigated the physiological response of participants to several HRI scenarios that involved playing games and interacting with robots. In one such scenario [73], participants played a Jenga game with either a human or a robot partner, and the conclusion of the study was that the participants exhibited more GSR peaks when playing the game with a robot compared to playing it to a human. In the other study, the participants played a modified version of the Stroop test [1] either alone or assisted by a robot. This study concluded that the participants exhibited again more GSR peaks and also blinked at significantly higher rate when playing the game in the presence of the robot compared to the control condition.

Hence, these studies show that humans exhibit different physiological manifestations when interacting with a robot compared to interacting with a human, at least in the experimental conditions defined by the researchers. This conclusion does not prove that the theories formulated with respect to the manifestations of deception in inter-human interactions are not valid in HRI, yet the burden of proof is shifted towards the hypothesis claiming that they are indeed valid and can be applied implicitly in HRI. Therefore, throughout this thesis, we will investigate these physiological manifestations without assuming that previous results from inter-human interactions are also valid in HRI, and we will also attempt to directly compare the physiological manifestations of participants taking part in deception detection scenarios in both inter-human and human-robot interactions.

2.6 Conclusions

In this chapter, we have analysed the results of past research done in the domain of deception detection, the vast majority of them being studied in inter-human interaction scenarios. We have discussed the two main approaches to detecting deception (the polygraph-based methods and the cognitive load based methods) and then detailed the major physiological and behavioural manifestations that have been proven

to be correlated to deceptive behaviours. Then, we have discussed the categories of measurement devices and techniques that have been used in order to evaluate these manifestations, classifying them according to their invasivity and comparing their effectiveness. Lastly, we have established that, at this point, there is no guarantee that humans exhibit the same manifestations and behaviours when deceiving a robot as when deceiving a human, and that the burden of proof lies with any theory claiming the manifestations are similar in the two scenarios.

As we have shown, each of the two main approaches to detecting deception has its flaws and limitations, as well as a number of countermeasures that can be employed in order to mislead the system and/or person evaluating one's trustworthiness. Moreover, the reliability of each technique and of the deception detection process in general is quite variable, from a detection accuracy of slightly over the 50% threshold attributed to chance to a very high accuracy of over 90% in some cases. On top of that, the reliability of some methods, particularly the cognitive load based ones, is strongly enhanced if appropriate interrogation procedures are used. Lastly, even if the polygraph-based methods have been used for a few decades already and have proven to be empirically successful in a significant number of cases, they have been largely contested by a part of the scientific community due to their lack of a solid theoretical background.

Even though we cannot assume these previous results obtained in inter-human interactions will also be validated by our research in HRI, they represent undoubtedly our source of inspiration and the starting point of our work. From the variety of manifestations and cues we have discussed, as well as their associated measurement devices and techniques, each with their strenghts and weaknesses, we will now have to choose those that we consider to be the most appropriate for our research. We need to take into account the nature and the limitations of HRI, focusing on our main goal, which is improving the ability of robots to detect deception in everyday interactions with humans and, in particular, improving their ability to assist human users who are in need of permanent and/or long term supervision and assistance. Hence, we generally favour the study of manifestations and cues that are not dependent on a particular interaction scenario nor of a highly invasive measurement technique, taking into account a series of experimental design guidelines that will be discussed in the next chapter.

Chapter 3

Experimental methodology and resources

3.1 Introduction

In order to test and validate our research questions, we need to design HRI experimental scenarios, during which we measure and analyse the physiological manifestations and behaviour of the participants. We have developed several experimental setups and we monitored several physiological manifestations, yet the resources and methodology we used were consistent throughout our work. In this chapter, we detail the common platforms, sensors, software, guidelines, and procedures we have used for our experiments.

This chapter defines the experimental design guidelines that have been established and the reasoning behind them, continues by presenting in detail the robotic platforms and sensors used for our experiments, the main software analysis tools and procedures used and follows up with the methods used to determine the psychological profile of the users, such as their personality.

3.2 Experimental design guidelines

The main objective of each of our experiments consists in enticing the participants to lie or deceive their interlocutor (i.e., robot or human), while being monitored by the sensors and devices used to evaluate their physiological response. In order to make sure the data we collect is indeed relevant, we followed a series of design guidelines:

- natural interactions
- exerting free will
- in-the-wild experiments
- truth must be known
- establish a baseline

Natural interactions

The experimental setups we designed involve, at some point, a discussion between the participant and their interlocutor, robot or human. This conversation between the participant and the interlocutor, especially when the interlocutor is a robot, needs to be as fluent and natural as possible. The interlocutor's discourse had to be clear and easy to understand by the participant, while the communication delays had to be reduced to a minimum, so as to reduce to a minimum any biases induced by the

dynamics of the interaction itself. Therefore, also due to the limited performance of the currently available robotic platforms and the software they come with, we decided to conduct *Wizard of Oz* experiments, where on one hand the robot's discourse is pre-recorded, using the experimenter's actual voice, and on the other hand the robot is remotely controlled by the experimenter throughout the entire experimental procedure. As a consequence, the robot's voice is clear and easy to understand, while the robot itself does not have to understand and react to the answers of the participants. However, the experimenter also has to be able to monitor in real time the interaction, either remotely or by being in the same room as the robot and the participant, but without interfering in any way in the interaction. As we wanted to ensure our experimental setups are as robust and less complex as possible from a technical point of view, we decided to keep the experimenter and the participant in the same room as the robot. This decision implies that we had to find ways to isolate the experimenter from the interaction while still allowing him to witness it and control the robot accordingly.

Exerting free will

Participants have to be enticed to lie or to deceive their interaction partner during one part of the interaction. The traditional approach to this problem is to offer to the participants rewards if they end up accomplishing a given objective, in this case successfully deceiving their interlocutor. However, in order to make the interaction as realistic as possible, the participants have to be able to exert their free will throughout the entire interaction and to reason themselves when and why they have to deceive their interlocutor. Therefore, they are not explicitly asked to tell the truth or lie when asked a particular question, but are promised rewards if they manage to accomplish a certain task that involves deceiving their interaction partner at some point. Even though the participants are **enticed** to lie by being promised rewards in exchange for some accomplishments that involved lying, they are free to choose at what point and especially how they end up deceiving their interlocutor and they are not coerced in any way.

Our post-experimental questionnaires, that the participants are asked to fill in once the experiment is over, also included a question concerning how strong their desire to earn the reward is. Its purpose was to evaluate to what extent the approach we use to entice the participants is effective or not and whether they are motivated to fulfil their goal or not.

In-the-wild experiments

Next, we decided to design our experimental setups to be as natural as possible (closer to an *in-the-wild* interaction). Therefore, we allow the participant to exhibit the maximum amount of free will and reduce to a minimum any imposed limitation or external control with respect to the participants' behaviour, as well as to their reasoning or decisions. Participants are not told to control their movements, posture, gaze, or voice in any way throughout the experiment, so as to make sure their behaviour and manifestations are as natural as possible. This is particularly challenging, as on one hand the more natural and unhindered a person's behaviour is, the more complicated to monitor and evaluate remotely it is, while on the other hand we have no guarantee the participants will actually perform as expected and lie or tell the truth at the expected moments.

Truth must be known

Another important aspect is to ensure that, when analysing each interaction, we are able to know exactly which of the participants' answers or statements are truthful and which are dishonest. This, of course, imposes some limitations concerning the scenarios we are able to design as well as the questions we are able to ask the participants. We therefore have a choice between questions whose correct answers are known to the participant *a priori* and questions whose correct answers are only known *a posteriori*, once a part of or the entire experiment is over. In both cases, we have to make sure we are able to establish post-experimentally whether the participants were honest or deceitful at each point of the interaction.

Establish a baseline

Last, but not least, each interaction has to begin with an initialisation or reference phase, during which the physiological manifestations of the participants are monitored and evaluated. The purpose of this phase is to establish a baseline for the parameters monitored for each of the participants. As we are interested in analysing the variations of the physiological and/or behavioural parameters and not their absolute values, establishing the baseline values for each participants is a key part of the data acquisition process that must be taken into account when designing the experimental setups.

3.3 Robotic platforms and sensors

A major choice we had to made in the beginning of this research thesis was the choice of robots and sensors used for our HRI experiments. One of the main guidelines of our experimental design was to make sure the interactions are as natural as possible, therefore we needed to use a robot that can perform this task as well as possible. Moreover, since another major objective of this research is to use only non-invasive and minimally-invasive sensors to analyse the participants' manifestations and behaviours, the choice of sensors was also crucial.

In the next subsection, the robot and the complete list of sensors used for all our research work are presented, together with their technical details and specifications and the output they provide. The exact way we have used the data they provided will be detailed in the following chapters, along with the physiological manifestations we monitored and the results we have obtained.

3.3.1 Robotic platforms

In this thesis we used the SoftBank Pepper robot. The physical appearance of the Pepper robot makes it appropriate for our research work. For instance, its height is suitable for interactions where the interlocutor sits down at a table in front of the robot, which were mainly the type of interactions we designed.

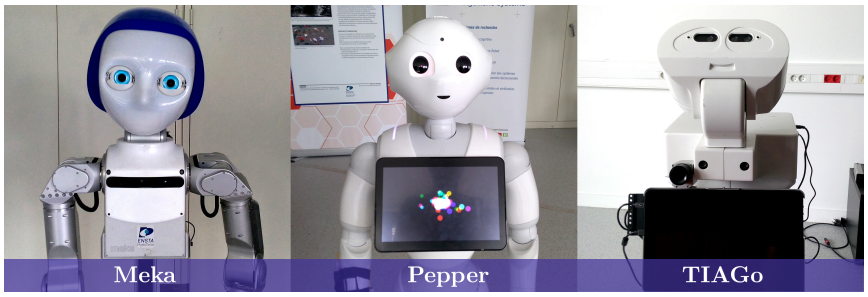

FIGURE 3.1: Meka, Pepper, and TIAGo robots from the Autonomous Systems and Robotics lab/U2IS

Some of the relevant technical specifications of the Pepper robot, as provided by SoftBank, are the following:

- **height:** 1.20 m (when standing up)
- **motherboard:** Atom Z350 1.6 GHz single-core CPU, 1 GB RAM
- **connectivity:** LAN + 802.11 a/g/b/n WLAN
- **cameras:** 2x RGB cameras (at the top and bottom of the head, 640x480 @ 30 Hz), 1x RGB-D Asus Xtion placed in the eyes (limited at 320x240 @ 20 Hz)
- **audio:** 2x speakers (on each side of the head), 4x microphones (on the head)

Even though the robot has integrated cameras (both RGB and RGB-D), we decided not to use them, due to their limited capabilities. In particular, the Asus Xtion RGB-D camera works in a limited performance mode, its resolution being limited to qVGA (320x240), while its frame rate is also reduced at 20 Hz. We therefore used a series of external cameras (i.e., RGB, RGB-D, or thermal) that we placed on fixed brackets in front of the robot.

Since the main task the robot has to perform is to speak as naturally as possible with its human interlocutors, its speech reproduction and speech recognition capabilities were also important. As mentioned in the first part of this chapter, the limited interaction performance of our robotic platform was the main reason behind our decision to implement *Wizard-of-Oz* experiments. In particular, the slow and somewhat unreliable speech recognition module that is part of Pepper's software meant that we had to rely to the experimenter to evaluate the participants' answers during the interactions and adapt the robot's speech accordingly. Moreover, since the robot's default voice was very similar to that of a child, we considered that particular voice tone and pitch is not adequate for our experimental designs. Therefore, we decided to pre-record the entire discourse using the experimenter's voice, both in English and in French. This guarantees that when comparing similar inter-human and human-robot interactions where the robot is replaced by the experimenter, there is no bias induced by the different voices or by any difficulty to understand the robot's speech.

3.3.2 Sensors and measurement devices

In this subsection, we will detail the specifications of the various sensors we have used for our experiments. The exact way they have been used will be detailed in the following chapters, when presenting the experimental setup and the physiological parameters we monitored.

RGB-D camera (Asus Xtion)

RGB-D cameras, of which the most popular model is the Microsoft Kinect, provide two video streams. The first stream is a classical RGB stream, displaying regular colour images of the objects placed in front of the camera. The second stream is the **depth** stream, that displays a representation of the distances between the camera and the points in space that can be observed in the RGB stream. More specifically, each pixel that can be viewed in the RGB frame is paired to a pixel from the Depth frame that contains the distance between the camera and the object depicted by that pixel. Based on this matrix of distances, a colour or grayscale representation of the depth image can be easily constructed, as well as a 3-dimensional visualisation of the area viewed by the camera.

FIGURE 3.2: The Asus Xtion RGB-D camera

We have used an Asus Xtion RGB-D camera, which similar in specifications to the first version of the Microsoft Kinect. The Asus camera provides **VGA (640x480)** outputs at a frequency of **30 Hz** for both the RGB and depth streams, but lower resolutions and frame rates are also available. The distance of use for the depth sensor is between 0.8m and 3.5m, any distances outside this range not being accurately measured. This camera can be connected to any computer using a regular USB 2.0 cable and is ROS compatible, using the **OpenNI2** ROS package that is available for several ROS distributions. Last, but not least, it has built-in microphones that allow us to record an audio stream during our experiments, using any audio recording software.

RGB camera (Logitech C920)

RGB video cameras are now commonly used in robotics and computer systems in general. However, the quality and resolution of the video stream they provide varies significantly from one model of camera to another, and especially in non-ideal lighting conditions. Because of these differences, we decide to also include a RGB camera that has a high resolution, noise-free image quality in less-than-perfect lighting conditions.

FIGURE 3.3: The Logitech C920 Webcam

In this work, we used the Logitech C920 webcam. It can provide a Full HD (1920x1080) output at a frame rate of 30 Hz, it uses USB 2.0 connectivity and its video stream can be acquired and published using the **usb_cam** ROS package. Similarly to the Asus RGB-D camera, the Logitech webcam has integrated microphones that allow us to record the audio stream during our experiments using standard audio recording software.

Thermal camera (Optris PI640)

Similarly to the way RGB video cameras provide a representation of the colours and lighting of objects placed in their field of view, thermal cameras provide a representation of the temperatures of objects placed in their field of view. These cameras are far less commonly used than RGB video cameras, especially in robotics or human-computer interaction applications, but several manufacturers provide cameras that have USB connectivity and are also ROS compatible, making them ideal for our applications.

FIGURE 3.4: The Optris PI640 thermal camera and an example of a thermal image representation

One such camera is the PI640 manufactured by Optris. Its sensor can measure temperatures between -20°C and 900°C, with a precision of 0.1°C. The sensor's resolution is 640x480, while the output frame rate is of 32Hz. The camera has USB 2.0 connectivity and is indeed ROS compatible. This ROS package provides two types of video streams, one containing the raw temperature data for each pixel of the thermal image, while the other one contains a visual colour-coded representation of the thermal image, that can be seen in Figure 3.4. The colour-coded representation was very useful for visualising the output, but we were more interested in the raw temperature values for our experiments.

Wireless sensor armband

In order to measure the participants' heart rate (HR) and galvanic skin response (GSR) in a non-invasive possible way, we have built a wireless and ROS compatible armband. This armband streams the values measured by its sensors in real time using a TCP/IP protocol to a server running on the computer overseeing the experiment, at a frequency of over 100 Hz, which then are published to individual ROS topics. The armband uses 2 sensors manufactured by Grove - a Grove finger-clip heart rate sensor and a Grove GSR sensor - as well as two analogue temperature sensors, one placed inside the armband core and the other placed inside the wrist band of the Grove GSR sensor.

The Grove HR sensor is a photoplethysmography (PPG) sensor that uses a green LED light and an optical sensor to measure the changes in the skin colour as blood flows through it periodically. This sensor computes the value of the estimated heart rate every second, using a number of previous measurements, and can be interfaced digitally using an I2C communication protocol. Even though the sensor is intended to be placed on a fingertip, it can be placed on any region of skin that is properly vascularised, such as the inside of the wrist or the inside of the elbow. Since we wanted to avoid hindering with the participants' dexterity during our experiments, we placed the HR sensor on the participants' left wrist.

The Grove GSR sensor is an analogical sensor that measures the skin conductance measured between two metal electrodes placed on the participants' fingers, using operational amplifiers. These operational amplifiers convert the value of the skin conductance between the two electrodes to a voltage of 0...5V. The sensor is therefore analogue and has a virtually null latency (or an extremely high measurement rate). However, in order to measure the analogue output, the use of an Analogue to Digital Converter (ADC) circuit is required. The two electrodes were placed on the participants' left index and pinky fingers.

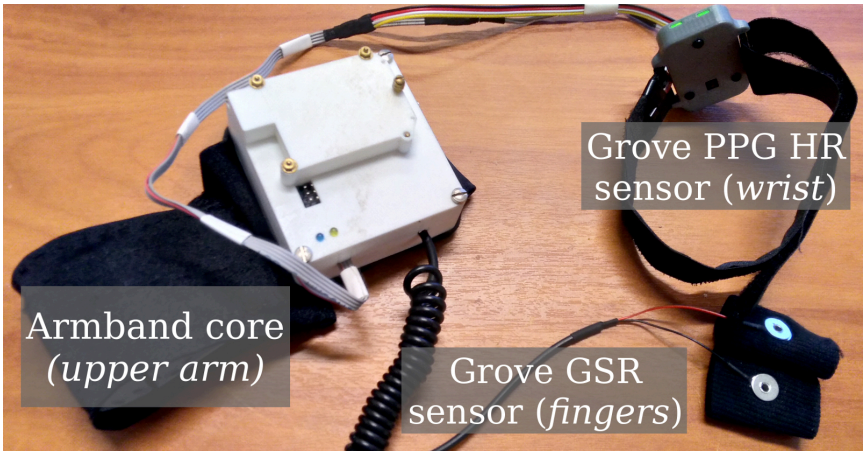

FIGURE 3.5: The wireless sensor armband

The core of the armband, which we developed ourselves, and that can be seen along with the two sensors in Figure 3.5 and in Figure 3.6, is built around a **NodeMCU ESP8266** development board. This open-source platform contains a 32-bit 80MHz micro-controller programmable using the Arduino environment (based on the avr-gcc libraries) and an embedded WiFi module with TCP/IP capabilities, along with a series of digital GPIO pins and one analogue input. This boards connects wirelessly to the TCP/IP server running on the PC overseeing the experiment and publishes

with a frequency of over 100 Hz the latest values provided by the sensors. The armband's core case is 3D printed, while the circuitry is powered by the Li-Po battery of an old camera, using a step-up 5V power regulator.

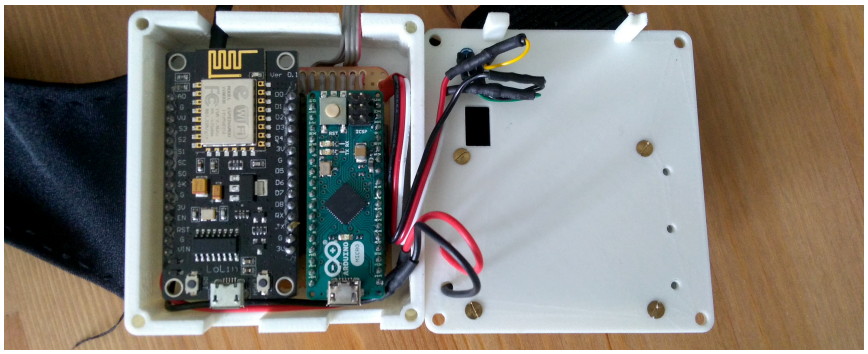

FIGURE 3.6: The circuits inside the wireless sensor armband

If the communication between the NodeMCU board and the Grove HR sensor is straightforward, using a digital I2C protocol, the use of analogue sensors is less obvious, since the development board has only one analogue input available. Therefore, we used a secondary board, more specifically an **Arduino Pro Mini** board whose ATMega328 micro-controller has an integrated 10-bit ADC with 6 analogue inputs. This micro-controller board was used to measure the values of all the analogue sensors (GSR and temperature), as well as the LiPo battery voltage, and then communicate them serially to the NodeMCU main board. Both the NodeMCU and the Arduino Pro Mini boards were programmed using the **Ino** command-line toolkit, which is a terminal-based compiler for code built using the Arduino libraries and the avr-gcc compiler.

On the computer server side, we have developed a ROS package that on one hand collects the data published periodically by the armband and on the other hand publishes the values of each sensor periodically to a series of ROS topics, one for each parameter (i.e., HR, GSR, wrist temperature, core temperature, battery voltage). These values can then be easily used by any ROS-compatible application or simply displayed in real time in a terminal. The TCP/IP communication was implemented using the SFML library's Network module.

3.4 Experimental monitoring and analysis software

The primary part of our work involved developing the software tools that would allow us to study the physiological manifestations associated to deception in HRI. Our software had to accomplish several objectives, some in real time and others after the experiments were over:

- fully control our *Wizard-of-Oz* experiments
- extract in real time a variety of physiological parameters
- record and replay all the raw experimental data for post-experimental analysis
- process the raw experimental data in order to obtain meaningful results

Moreover, we wanted to make sure that the software we develop reduces to a minimum the number of repetitive and time-consuming tasks that the experimenter has to do in order to prepare, undertake, and analyse the results of the experiments.

On top of that, we decided to use technologies and libraries that are open-source, free to use, and also as platform-independent as possible, so that not only the results, but also the software obtained throughout our work could be later re-used or improved.

Therefore, we decided to create two distinct software components. The first one is an application that oversees the experiments. It controls the robot, acquires and records the raw data from all the sources (video cameras, microphones, various sensors), but also is able to replay the prerecorded raw data. Moreover, this application performs all the raw data processing necessary to extract any physiological parameters that we attempt to monitor in real time, either during the experiment or afterwards, based on the recorded sensor data. However, this application is supposed to deal only with one experimental run at a time, which means that all post-experimental analysis has to be done externally.

Hence, our second main software component is the one dealing with the post-experimental analysis of the data collected from all participants to an experiment. Moreover, in the case of physiological parameters that cannot be estimated in real time, but need an analysis on longer time intervals, this component is also responsible of extracting those parameters. Since this software component is supposed to process large amounts of data containing physiological parameters collected from several participants, we also wanted to automate the batch processing to the highest possible extent. Lastly, the purpose of this second software component is to find out if there are any correlations between the parameters we monitored and the experimental conditions.

In this section, we will discuss the architectures, methods, and libraries used for each of these two major software components, the functionality they provide, as well as the output they each produce.

3.4.1 Experimental software application

We started by identifying the needs our main software application has to fulfil. Firstly, it communicates with all the hardware components involved in our experimental setups. It controls the robot interacting with the participants, it acquires data from the video cameras (RGB, RGB-D or thermal) used in our experimental setups, as well as from the cameras' microphones, and also recovers the data provided by other sensors (such as, for example, the wireless sensor bracelet measuring the participants' HR and GSR). Moreover, since we want to be able to review the raw experimental data at future times, as well as perhaps to change or update various physiological parameter extraction algorithms, this application also has to be able to record and replay all experimental raw data. Last, but not least, we wanted the application to have an easy to use interface, and also to be as efficient when it came to the use of computer resources, as it had to run on a regular PC.

Software tools and resources

The language of choice for the development of this software component was C++. We developed and ran it using Qt Creator 4.0.2, under the Linux Mint 17.3 operating system, on a Lenovo ThinkPad T430 laptop computer. The main reason for using C++ was the improved efficiency and execution speed compared to other languages, especially when all available compiler optimisation flags are used, as well as the fact that it allowed us to use open source libraries and tools that allowed us to interact with all the hardware components of our system. The only other language allowing us to use the same hardware and software tools was Python, but C++ was preferred

because of its superior performance. We relied heavily on the use of several software libraries: various ROS Indigo packages to recover data from the cameras and sensors we have used, but also to record and replay the raw experimental data, the SFML 2.4.1 open-source library for the graphical user interface design, the 2.3.3.28 version of the NAOqi SDK for controlling the SoftBank Pepper robot we used for our HRI scenarios, as well as the OpenCV 2.4.11 library for various image processing tasks. Moreover, we used two different face tracking libraries, the CLM Face Tracker [60] developed by Tadas Baltrusaitis, as well as the face tracking module of the DLib C++ library [62].

The graphical user interface

The GUI of our software, developed mainly using the SFML C++ library, allows the experimenter to easily perform the following tasks:

- visualise the output of several video cameras, depending on the experimental setup
- control the activity of the face tracking module (start, stop, reset)
- visualise the activity of the face tracker and the identified face
- visualise a series of physiological parameters (both text and graphs, depending on the experimental setup)
- record and replay experimental data, with a file chooser pop-up list for the replay
- connect the Pepper robot and test the connection
- step by step control of the experimental procedure
- customise or validate experimental conditions, depending on the experimental setup
- view the status and output of other sensors

Even though the interface cannot be customised on-the-fly, its modular architecture allows us easily to add or remove buttons, to add, remove or reconfigure video camera monitors or to add text output for more physiological parameters. Rather predictably, its design changed slightly from one experimental setup to another, but the functionality it offered remained the same. The following picture shows the final version of our GUI, used for our last experimental setup:

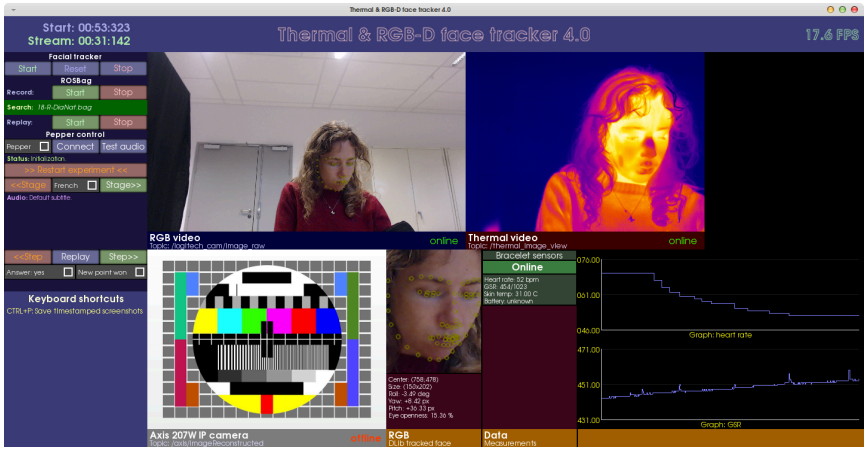

FIGURE 3.7: The final version of our experimental GUI

Controlling and monitoring the experiments

One of the key guidelines of our experimental design was the development of *Wizard-of-Oz* experiments, which implies controlling the robot remotely throughout the interaction it has with a human interlocutor. In particular, we wanted to be able to easily play the experimental discourse and questions that the robot asks during an HRI scenario, as well as to dynamically adapt the robot's discourse to the answers given by the participant to some key questions. This dynamic adaptation had to be made quickly that in order to avoid perceptible delay in the dialogue between the human and its robot interlocutor, meaning that the interface had to allow us to do this easily.

Therefore, we designed our experimental discourses as a series of statements or questions, several of them being grouped into an experimental **stage**, while each being considered an individual **step** of the discourse. In the case of simple statements, such as for example explaining the experimental rules to the participant, or questions whose answers do not have an impact on the choice of future questions, the experimenter overseeing the procedure only has to sequentially advance from one step to another. Advancing from one step to the other is done instantly, so no interaction delay is added.

However, in the case of any sort of interrogation where the participants' answers are relevant, the experimenter also has to manually register the answer given by the participant, based on the answer that the participant has just given, using one or several check-boxes. This has to be done as quickly and accurately as possible, so as the interaction does not suffer from noticeable delays. A similar mechanism of check-boxes is used, for example, to decide whether the participant ends up receiving their reward or not, or for any other adaptation of the experimental discourse to conditions that vary during the experiment.

Moreover, in case the participant does not understand one question or avoids the answer to a particular question, the application allows the robot to easily repeat the question or even go back to a previous question. Last, but not least, since some participants had a preference for the language used by the robot, we made sure that the application can easily switch between English and French for the robot's experimental discourse.

The structure of each experimental discourse, with all the potential variations based on the participants' answers, is coded as a Finite State Machine, with the experimenter advancing manually from one step to another, either using the only logically possible transition or automatically deciding the appropriate next step based on the participants' statements or other external conditions. Each experimental setup involves creating and testing a different Finite State machine, that makes sure that all possible scenarios are taken into account and verified before the experiments actually start.

Using ROS to acquire sensor data

The backbone of our software uses ROS to connect to the various cameras and sensors used for our experiment. Each ROS package then publishes the data collected from a camera or sensor to one or several ROS topics, to which the main software application then subscribes. Once the main application subscribes to this series of topics, every time new sensor data is available, the application will automatically receive the input data from the corresponding ROS topic and process and display it accordingly.

Recording and replaying experiments

Another functionality that our GUI provides is the ability to record and replay the exact raw data provided by the ROS-compatible sensors during an experiment. This has been done using the **rosbag** ROS package, that allows the recording of the data that is being published to any number of ROS topics. We have integrated the functionality provided by this package in our main software application, using a series of BASH scripts that are executed by the application, also allowing the user to manually choose between several prerecorded files using a pop-up list.

However, the audio input provided by the microphones of the cameras used during our experiments was not processed by ROS packages, we had to record it directly as an audio file. For this task, we have used the Audacity open-source software, that recorded the audio and exported .mp3 files. These files were then manually synchronised with the video output provided by the cameras and replayed in tandem with the **rosbag** files when replaying experimental data in the main software applications. Moreover, these audio files were crucial for the data annotation process that will be detailed later in this chapter.

Extracting physiological parameters

The final main task fulfilled by the main software application is extracting a series of physiological parameters based on the input provided by the various sensors. Some of the parameters can be extracted on-the-fly, running a series of algorithms based on the last input provided by the sensors, while others can only be extracted post-experimentally, which is the case for all parameters that are evaluated using frequency-domain analysis, described in the following section. In the latter case, the main software application will instead extract at each frame a preliminary measure that will be later analysed and post-processed by the post-experimental analysis algorithms.

Nonetheless, for each of the extracted parameters or preliminary measures, the application has the same work flow. For each new frame or data package provided by the sensors used (cameras or wearable sensors), the freshly received data is processed by the series of algorithms extracting the various parameters or preliminary measures, and then the result of each of them is logged to a timestamped CSV file. Each timestamped CSV file contains the raw value extracted for each parameter and the relative time with respect to the moment the application was launched, with a precision of 10^{-6} seconds. Moreover, some of these parameters are also displayed on the GUI, so the experimenter can monitor them during the experiment on when replaying the experimental data.

These parameters can be computed both during the experiment and also when replaying the raw experimental data previously recorded. Since we wanted to reduce to a minimum the CPU load during the experiment, we therefore decided to only record the raw experimental data during the experiments and then replay every experimental data set once and run all the physiological parameter extracting algorithms on the prerecorded data. This redundancy allowed us on one hand to record the experimental data at the highest possible frame rate and also to monitor the extraction process via the GUI after the experiment was over and the experimenter did not have to oversee the experiment and the HRI anymore. Nonetheless, since all the algorithms in question could have also been run in real time during the experiment, we do not consider this to be a part of the post-experimental data analysis.

3.4.2 Post-experimental data analysis

Regardless of the experimental setup or the physiological or behavioural parameters we monitor, the analysis procedure we have implemented follows the same routine. The main software overseeing the experiment exports, for each monitored parameter and for each participant, a timestamped CSV file. Each of these files contains the raw values of the extracted parameter, as well as a relative timestamp with respect to the start of the experiment, as the sampling frequency is variable and depends on the CPU load during the raw data extraction phase.

This raw, timestamped data needs to undergo several phases of processing and analysis, that can be separated into two major stages:

- transforming the raw data into meaningful measures
- finding statistically significant correlations

From raw data to meaningful measures

This raw, timestamped data, exported by our main software application by means of CSV files, is processed using Scilab scripts so that it can later be used to search for statistical correlations between the experimental conditions (i.e.: truth value of the participants' statements, nature of the interlocutor, various personality traits) and the evolution of the various monitored parameters. Depending on the type of analysis we make (time-domain or frequency-domain), some intermediate steps are taken differently, but the main approach for data analysis, that will be detailed in this section, is the following:

1. Manually annotate the experimental data (timestamps and truth values of their statements).
2. Segment the raw data based on the time annotations.
3. Process the raw data on each segment (**time-domain** and/or **frequency-domain** analysis, depending on the monitored parameter).
4. Create summary files containing the results of the data processing for each interval, as well as the truth value associated to them.

Annotations and data segmentation

Using the recorded experimental data (video rosbag files, as well as audio files), the experimenter has to manually annotate a series of relevant elements of information. First of all, we have to identify the beginning and the end timestamps of each relevant experimental interval. Regardless of the experimental setup, the first significant interval is the **baseline or reference interval**. Then, depending on the experimental setup used, we identified several analysis intervals, with a timestamping precision of 1 second. We will detail the actual number and structure of the intervals when presenting each experimental setup.

The other purpose of the annotation process is evaluating the truth value of the participants' statements. Since the experiments were designed in such a way that the true answer of each question is known by the experimenter at analysis time (or even beforehand, depending on the questions), it is easy to evaluate the truth value of each answer or statement. Therefore, when annotating the data, the experimenter also establishes whether the answer of each question asked by the participants is true, false, or inconclusive. Then, depending on the experimental setup and the way

we define the analysis intervals, the experimenter will establish the overall veracity of the participants' answers throughout each interval based on the veracity of the answers of all the questions answered during that interval. The exact procedure for establishing whether a participant was deceptive or honest during an interval will also be detailed when presenting each experimental setup, in the next chapters.

Time domain analysis

Depending on the physiological variables we monitor, we distinguish between two major types of analysis: time-domain and frequency-domain analysis. In most of the cases, the time-domain analysis is preferred, as we are interested in analysing the long-term evolution of variables and not in identifying periodical events or variations and extracting the frequencies of such events.

Regardless of the parameter, the time domain analysis procedure is identical. For each interval, including the baseline interval, we compute the average value, the standard deviation, and the slope of the linear interpolation of the recorded samples. Then, for each interval except the baseline interval, we compute the relative average value, relative standard deviation, and relative slope of the linear interpolation with respect to the baseline interval. At the end, the relative aforementioned measures are each saved in individual CSV files, where on each line we store the relative value for the given interval and the truth value associated to the given interval, based on the manual annotations. These files are later used to establish statistical correlations between the parameter evolution and the truth value associated to each interval.

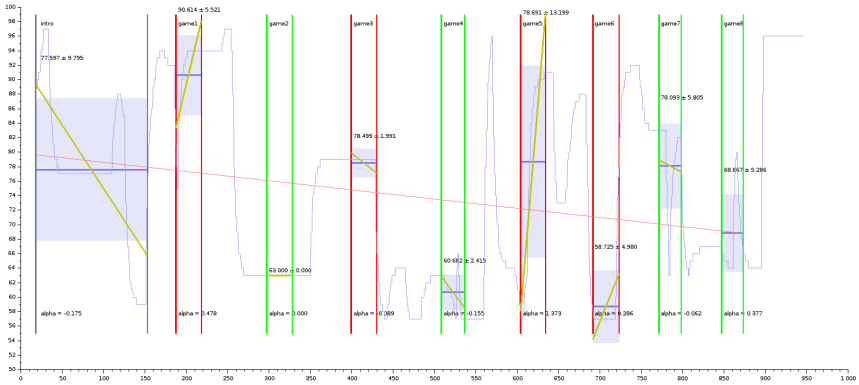

FIGURE 3.8: The evolution and the processing of a participant's HR

Moreover, the analysis procedure allows us to visualise the evolution of a given parameter throughout the experiment, as well as the results of the processing. The graph in Figure 3.8 shows us an example of the evolution of a participant's heart rate (HR) during one of our experiments, as well as the result of the annotation process, and the computed measures. The colour-coded vertical lines separate the analysis intervals: grey lines for the baseline interval, green for the intervals where they were honest, and red lines for the intervals where they were deceitful, while the thick lines signify the beginning of an interval and the thin ones its end. The light blue graph displays the raw data, with time measured in seconds on the horizontal axis and the HR measured in BPM on the vertical one. The thick blue horizontal lines and the light blue rectangles show the average value and the standard deviations for the given interval (also displayed numerically), while the yellow lines show the linear interpolation of the data over each interval, with the slope (alpha) also being numerically displayed for each of the intervals.

Frequency domain analysis

Some of the physiological parameters that we analysed, particularly those whose evolution was periodical, require a different type of processing. Instead of analysing their long-term evolution, which implies a time domain analysis, we needed to extract the frequency of the periodic evolution of those parameters. This requires a FFT analysis of the raw data samples exported by the main experimental application, that allows us to identify the frequencies of the most important periodic variations of the data samples.

Therefore, after segmenting the raw data using the same segmentation procedure described earlier, we start by interpolating the raw samples at a fixed frequency of 30 Hz, in order to compensate for the variable logging frequency, and then perform a FFT analysis on each relevant interval, including the baseline interval. Then, depending on the physiological parameter that we monitor and the band of frequencies inside of which we expect to find it, we filter the resulting spectrum and only select the frequencies that could be considered physiologically realistic. Last, we select the 3 most intense peaks from the filtered spectrum and the relative intensities of the second and third most intense with respect to the most intense. We also compute a weighted mean frequency, using the relative intensities, over the three most significant frequencies.

The reason for extracting the 3 most intense peaks from the filtered spectrum, as well as for computing the weighted mean frequency, is that the parameter whose evolution we are studying does not necessarily have a constant frequency over the measurement interval. In the particular scenario where the given parameter has the same exact frequency over the entire interval, then the most significant frequency would indeed correspond to the real frequency we are searching for. However, since we could not make this assumption, the most significant frequency would then correspond to the **median** frequency of the parameter in question. Therefore, extracting several most significant peaks as well as their relative intensities with respect to the most significant one allows us to take into account to some extent the variations of this frequency over the sampling interval.

Lastly, in a similar fashion to the procedure for the time domain analysis, we computed the relative values of the extracted frequencies with respect to the baseline interval, and then exported the relative frequencies (most significant 3 and the weighted mean frequency) to individual CSV files, where on each line we store the relative frequency of each interval with respect to the baseline interval and the truth value associated to that interval.

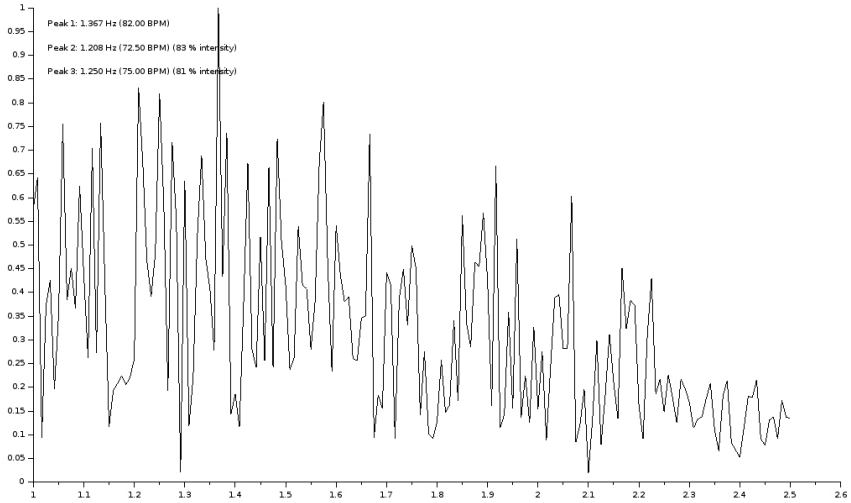

FIGURE 3.9: FFT analysis of the facial green colour component

Also similarly to the time domain analysis, we are also able to visualise the output of the FFT processing. The graph in Figure 3.9 shows the frequency spectrum generated by the FFT analysis of the evolution of the average facial green colour component over a given interval, filtered over frequencies in the 1...2.5 Hz interval. This analysis is done in order to estimate the participants' HR based on an RGB video recording and, in theory, the frequency of the most significant frequency peak should correspond to the median HR over that interval. This procedure will be further detailed in chapter 5.

Finding statistically significant correlations

The main purpose of our research, as already stated, is to search for statistically significant correlations between the evolution of participants' physiological or behavioural parameters and the veracity of their answers. Moreover, we are interested in using a similar approach to understand if there are any statistically significant differences between the participants' manifestations when taking into account other experimental conditions, such as the interlocutor's nature (human or robot), or one of the components of the personality profile of the participants.

In order to do so, we have to analyse the data obtained after the first analysis stage, which involves processing the raw data obtained from each participant during an experiment, using standardised statistical tests. The structure of our processing algorithm, fully automated for easy batch processing regardless of the number of parameters, participants and data samples, and implemented in Scilab and MATLAB, is the following:

1. **Merge and classify:** For each measure of each parameter, merge the data from all the participants. If needed, classify the merged data into several subcategories, depending on experimental conditions other than the truth value of the participants' statements (i.e.: human interrogator vs. robot interrogator conditions).

2. **Evaluate the data sets:** Use statistic tests to establish if the data in each merged data set is normally distributed or not.

3. **Search for correlations:** Depending on whether a data set is normally distributed or not, use standardised statistic tests to evaluate if there is a statistically significant correlation between the chosen condition and the values of the chosen measure of a given parameter.

Merge and classify

Regardless of the type of analysis previously done (in the time domain or the frequency domain), we obtain a set of CSV files. Each CSV file corresponds to a specific measure of a given parameter for one participant to an experiment and contains a series of lines where the relative measure during a key interval of the experiment with respect to the baseline interval is paired with the established truth value of the participant's statements during that interval. The first step of the statistical analysis process is to merge the data obtained from all the participants into one or several data sets, depending on the experimental conditions.

Each data set contains the specific relative measure of a given parameter for one experimental interval paired to the class it has been associated with. For example, when searching for correlations between the evolution of a physiological parameter and the veracity of the participants' answers, the classification is done based on the veracity of the answers, therefore having two classes, one for true answers and the other for lies. However, when also taking into account another experimental condition, such as for example **the nature of the interlocutor**, we generate four data sets, each with two classes:

- **truth vs. lie:** data set containing all the data samples separated into two classes, depending on the veracity of the answers, but without taking into account the nature of the interlocutor
- **robot vs. human:** data set containing all the data samples separated into two classes, depending on the nature of the interlocutor (robot or human), but without taking into account the veracity of the answers
- **truth vs. lie, robot only:** data set containing only data samples from the interactions undergone with a robot interlocutor, separated into two classes, depending on the veracity of the answers
- **truth vs. lie, human only:** data set containing only data samples from the interactions undergone with a human interlocutor, separated into two classes, depending on the veracity of the answers

This preliminary classification then allows us to analyse separately the impact of each experimental condition. Similar classifications to the previous example were done by taking into account various components of the participants' personality profile.

Searching for correlations

Before actually searching for statistical correlations, we have to make sure the data sets are eligible for the chosen standardised statistical tests. In particular, we distinguish between **non-parametric** tests and **parametric** tests. Non-parametric tests, such as Kruskal-Wallis test, do not require the data to be normally distributed for it to be eligible for the test, while parametric tests, such as the ANOVA test, require the data to be normally distributed. Moreover, other tests, such as the t-test, require the samples in each of the two classes to have equal variances. Therefore, before using any statistical tests, we have to verify if the data in each data set is normally distributed or not using the **Shapiro-Wilk** test, while the normally distributed data sets are also tested using the **Two-sample F-test for equal variances**.

Finally, depending on the properties of the data sets that we have previously established, we check if the null hypothesis (the assumption that in each group we

have random samples from the same population, therefore proving there is no statistical correlation between the membership to one of the two classes and the monitored parameter) is confirmed or rejected using one or two of the following tests:

- **ANOVA:** for data sets with a normal distribution
- **t-test:** for data sets with a normal distribution and classes having equal variances
- **Kruskal-Wallis**: for data sets that did not have a normal distribution
- **Wilcoxon rank sum test**: for data sets that did not have a normal distribution

If the null hypothesis is confirmed by all the tests (which means that the resulting *p-value* was higher than 0.05), we consider there is no statistically significant correlation between the evolution of the chosen measure of the given parameter and the membership to one of the two classes (true statements vs. lies, robot interlocutor vs. human interlocutor, etc.). If, on the other hand, at least one test rejects the null hypothesis, then we consider a statistically significant correlation was found. This part of the analysis is done in MATLAB, as it offers a large variety of dedicated functions that implement standardised statistical tests.

3.5 Personality profiling

The participants' various personality traits may have an impact on their behaviour during the experiments, as they do on everyday inter-human interactions. Therefore, analysing and understanding the personality profiles of our participants is crucial for two reasons. Firstly, because we need to make sure that there is no personality-induced bias in our experimental groups by verifying their distribution according to each personality trait is as even as possible. This is particularly important in the cases where the participants are separated into two (or several) experimental subgroups depending on a particular experimental condition (such as, for example, interacting with a robot interlocutor or a human interlocutor). Secondly, we need to evaluate their personality because we want to check to what extent their personality traits have an impact on the statistical correlations between their behavioural or physiological manifestations and deception.

To achieve this twofold objective, we have studied the literature in order to find out what are the commonly used standardised tests for personality profiling and which various personality traits they monitor. In particular, we have decided to use the BIG5 Personality Test [86] and the Eysenck Personality Questionnaire [87]. Our choice of the BIG5 Personality Test was motivated by the fact that it is one of the most used personality tests in literature, while the main reason behind the choice of the Eysenck personality test was its evaluation of the Lie dimension of the personality. This dimension is of particular interest to our research, as it evaluates the participants' predilection to lie, and the Eysenck Personality Questionnaire is one of the first and the few to evaluate this trait.

These tests each compute a score for each of the monitored personality traits. Then, using each of the score, we can either analyse the overall distribution of each personality trait within each of our experimental groups, making sure the distribution is as even as possible in case of several groups. Another way to use these results is to separate each experimental group into two subgroups, one with participants exhibiting high levels (higher or equal to 3) and the other with participants exhibiting a low level (lower than 3) of each of the personality traits.

BIG5 Personality Test

This test, developed by Goldberg in 1992 [86], allows us to evaluate 5 major personality traits:

- Openness
- Conscientiousness
- Extroversion
- Agreeableness
- Neuroticism

The Openness, or openness to experience, is defined by the appreciation of adventure, imaginative ideas, curiosity and variety of experiences the person is willing to try. Conscientiousness is highly related to self-discipline, organisation and a sense of duty. Extroversion is characterised as the willingness to engage with the external world, to interact with people, to have a behaviour that is full of energy and enthusiastic and to communicate a lot. Agreeableness reflects a person's ability to get along with other people and compromise easily when ending up in conflictual situations. Last, but not least, Neuroticism is linked to a predilection to experience negative emotions and also to be emotionally unstable, reacting more easily and negatively to stress or other external factors.

The test is comprised of 45 questions regarding the participants' own perception of themselves in various hypothetical situations and scenarios. The participants are asked to evaluate how they consider themselves with respect to the given scenario on a 5-point Likert scale, with 1 corresponding to "Strongly disagree" and 5 to "Strongly agree" with each of the hypothesis. The test algorithm then computes a score from 1 to 5 for each of the 5 aforementioned personality traits, with 1 corresponding to a low level and 5 to a high level. The complete question list of the BIG5 Personality Test can be found in Appendix A.

Eysenck Personality Questionnaire

Another popular standardised test is the Eysenck Personality Questionnaire [87]. This questionnaire has several versions, with varying numbers of questions and formulas used to evaluate the personality traits based on the participants answers, but we decided to use its short, revised version, comprised of 48 questions that have to be answered by "Yes" or "No". The questions are related to a variety of hypothetical scenarios and the participants have to answer what they consider they would or usually do in each situation.

This personality test evaluates 4 personality traits, two of them also being evaluated by the BIG5 Personality Test:

- Extroversion
- Neuroticism
- Psychoticism
- the predisposition to Lie

We were particularly interested by the fourth dimension of the EPQ questionnaire, the EPQ-L or Lie dimension, as it directly quantifies to what extent a participant is willing to lie during our game. This was also the main reason for which we

decided to use this test as an alternative to the BIG5 Personality Test. The complete question list of the short, revised version of the Eysenck Personality Questionnaire can be found in Appendix B.

3.6 Summary

In this chapter, we have discussed the exhaustive list of hardware and software experimental resources we have used throughout our entire research, as well as the methodology used to design our experiments and to analyse the data. The guidelines and data analysis methods are used for the design of every single experiment, while the hardware varies from one experiment to another, depending on the parameters we decide to measure and the measurement techniques we use. Moreover, we have also discussed the methods used to analyse the participants' personality profile. The purpose and use of each personality test is detailed in the case of each experiment in the following chapters.

3.7 My contribution

The contributions to this chapter consist of the list of experimental design guidelines, all the software developed for acquiring and analysing the data (with the exception of the external libraries and packages that we have mentioned), both for the main software application and the Scilab / MATLAB scripts used for post-experimental analysis, the design of the graphical user interface, as well as the Finite State Machine approach for the control of the robot during the HRI experiments. Moreover, we have also built the wireless sensor armband used to monitor the HR and GSR and we equally developed its embedded software, as well as the ROS package used to interface it with the computer.

Chapter 4

Detecting deception using only RGB-D and thermal cameras

4.1 Introduction

The first and major objective of our research work is to explore noninvasive lie detection techniques in HRI. In particular, we are interested in the use of RGB-D as well as thermal cameras as primary ways of detecting deception, using state of the art measurement techniques that have been discussed in chapter 3 and that have already been used in other HRI scenarios or in inter-human interaction scenarios. We have started by designing an experimental setup that would follow the guidelines discussed in section 3.2. Then, we focused on developing and implementing a set of measurement techniques that use solely a RGB-D camera and a thermal camera as measurement devices, trying to evaluate a series of physiological parameters that have been proven to be correlated to deception in inter-human interactions, as discussed in chapter 2.

In particular, our objective is to find out if there are any correlations between the participants' **interaction distance**, **eye openness**, **estimated heart rate**, **estimated respiratory rate**, and the **relative temperature of the inter-ocular and nasal areas** with respect to the face, in a **mock-up crime and interrogation** experimental scenario. In this chapter, we start by detailing the experimental setup we used, then continue with our attempts to detect deception using RGB-D and thermal cameras, as well as our results and the problems and limitations we came across, as well as of a description and analysis of results of the post-experimental questionnaire we used to evaluate our experiment. We will finish by summarising to what extent these techniques can be used in order to detect deception in HRI.

4.2 Experimental setup: mock-up crime and interrogation

The experimental setup that we have designed has two parts, that take place in separate rooms: a **mock-up crime** and an **interrogation** during which the participants are asked several questions concerning what they have done during the **mock-up crime** part of the experiment. This experimental setup had two versions, with some minor modifications made in order to improve the interrogation procedure. What we will discuss in detail is its **final version**, but we will also mention the changes that have been made with respect to the initial one, as well as the reasons we had to make them.

4.2.1 The mock-up crime

The participants are briefed about the experiment, instructed that they can stop and leave the experiment at any moment, and then they sign the informed consent form. Once they are ready to start, they are told to enter the first experimental room, where they find a sheet of paper with detailed instructions placed on a table. Alongside the sheet of paper, they find a chocolate bar, which is their reward, and five objects: a red marker, a computer mouse, a 20€ bill, a book, and a mobile phone. They are instructed to start by locking the door behind them. Then, they have to choose 3 out of the 5 objects placed on the table and hide them wherever they want inside the room. They are also instructed that they can take the chocolate bar with them when leaving the room, but that they will be allowed to keep it **only if they do not disclose any information concerning the objects they have hidden** inside the room. Finally, they are instructed to unlock the door, leave the current room and then go straight to the interrogation room.

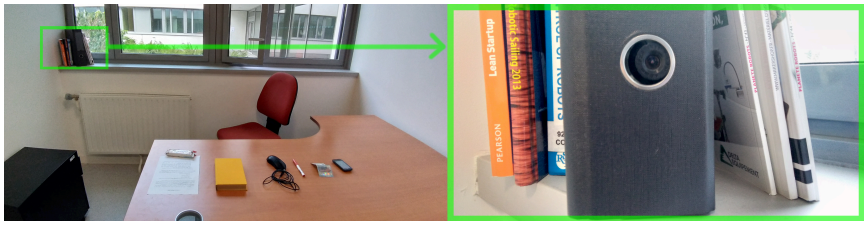

FIGURE 4.1: The mock-up crime room: the table, the objects and the hidden camera

During this phase of the experiment, their activity inside the room is monitored using a **hidden wireless camera**, that allows the experimenter to know exactly what objects they have hidden before they entered the interrogation room and adjust the interrogation discourse and questions list accordingly. Moreover, this information is crucial in order to establish, at the end of the experiment, whether the participants are allowed to keep their reward or not.

4.2.2 The interrogation

The interrogation phase is as follows. The participant enters the interrogation room, where a robot interrogator (a Softbank Pepper robot) sits behind a table, with an Optris PI640 thermal camera and an Asus Xtion RGB-D camera mounted on a photography tripod in front of the robot. No other robots or humans can be seen inside the room. The experimenter supervising the interrogation and controlling the robot is hidden behind some black curtains. The participant is invited by the interrogator to take a seat on the other side of the table and then the latter proceeds to explaining the interrogation procedure. The participant is told that they will be asked a series of questions concerning what they have just done during the first phase of the experiment, as well as some questions that have nothing to do with it. The interrogator also tells the participants that they are free to choose whether they tell the truth or they lie during the interrogation.

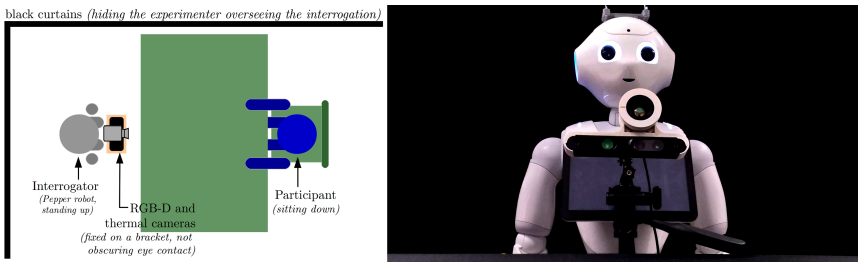

FIGURE 4.2: The interrogation room setup and the Pepper robot

Then, the actual interrogation begins and the participant is asked 4 different series of questions:

1. General questions, not related to the experiment:
   (a) What is your full name?
   (b) Are you sitting down right now?
   (c) What is your gender?
   (d) Are you wearing glasses right now?
   (e) What day of the week is it today?
   (f) What date is it today?
   (g) What colour is your hair?
   (h) Am I a robot?
   (i) Are you over 18 years old?
   (j) Where are you right now?
   (k) Have you been completely honest during this interview?

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):
   (a) Were you in the room number R231 just before coming here?
   (b) Did you see a table in the room number R231?
   (c) Did you see several objects placed on the table inside the room number R231?
   (d) Did you see a marker on the table?
   (e) Did you see a 20€ bill on the table?
   (f) Did you see a book on the table?
   (g) Did you see a computer mouse on the table?
   (h) Did you see a mobile phone on the table?
   (i) Did you do anything with [object number 1, not hidden]?
   (j) Did you do anything with [object number 2, not hidden]?
   (k) Have you been completely honest during this interview?

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).
   (a) Did you do anything with [object number 3, hidden]?
   (b) you do anything with [object number 4, hidden]?
   (c) Did you do anything with [object number 5, hidden]?
   (d) The [object number 3, hidden] is not on the table any more. Did you hide it?
   (e) The [object number 4, hidden] is not on the table any more. Did you hide it?

(f) The [object number 5, hidden] is not on the table any more. Did you hide it?
(g) Where did you hide the [object number 3, hidden]?
(h) Where did you hide the [object number 4, hidden]?
(i) Where did you hide the [object number 5, hidden]?
(j) Have you been completely honest during this interview?
(k) Do you think you deserve to keep the chocolate bar that you have just received?

4. The same exact general questions, not related to the experiment, asked in the first phase of the interrogation.

Finally, right after the last question series is over, the participants are informed by the Pepper robot if they are allowed to keep their reward or not, based on the objects they have hidden and whether they disclosed any information concerning what they have done with the objects in question. This decision is made, of course, by the experimenter that controlled the robot and oversees the experiment, using the video footage provided by the hidden camera in the mock-up crime room and the participants' answers during the interrogation. If a participant has disclosed any information concerning what they have done with any of the hidden objects inside the first experimental room, they are not allowed to keep their reward. The participants are then notified that the experiment is over and invited to leave the room. Afterwards, they are asked to fill in the post-experimental questionnaires.

Differences with respect to the initial version

The main differences between the first and the final mock-up crime setups are the fact that the number of objects to be hidden was changed from 2 to 3 out of 5, therefore allowing the interrogator to ask more questions concerning the hidden objects. Besides that, in the first version of the experiment, after the initial, general questions were asked, the questions concerning the hidden objects and those not related to the hidden objects were mixed together in a random order. Therefore, in the first version we only had a single interrogation phase of 15 questions, that contained the relevant, mock-up crime related questions, along with some more questions unrelated to the experiment. The main reason behind these changes was the need for larger analysis intervals. Therefore, we decided to segment the physiological manifestations of the participants on an entire question series basis instead of an individual question basis. The initial, question-based, analysis intervals were too short, the consequence being that the results were inconclusive. On top of that, the main reason for increasing the length of the analysis intervals was the need of a larger than 30 seconds window for the FFT analysis done in order to estimate the heart and respiratory rates with a good accuracy. We will detail later in this chapter all the reasons why the initial setup was not appropriate.

Last, but not least, in the initial version of the experiment the interrogation was conducted by a human experimenter instead of a robot, as we were interested in studying the differences between the participants' manifestations when interrogated by a human and those exhibited when being interrogated by a robot. At the end, we decided to only focus on the robot interrogator condition.

4.2.3 Annotating the experimental data

First of all, as part of the annotation procedure, we identified the beginning and the end timestamps of each relevant experimental interval. In the final version of the

mock-up crime and interrogation scenario, we identified 4 analysis intervals, one for each of the 4 question series, the first of the four being the **reference or baseline interval**. In the end, the final interval (containing the same questions as the baseline interval) was not included in the analysis process. Each interval started when the interrogator started asking the first question of the series and ended just after the participant finished answering the last question of the series and the annotation was done with a precision of one second.

Nonetheless, as we abandoned the initial answer-based time segmentation approach, we also had to adapt the way we evaluated the veracity of the participants' statements, which implied establishing the overall veracity of the participants' statements during each of the analysis intervals. Therefore, for the final version of the mock-up crime experimental setup, we decided to count the number of true, false, and inconclusive answers given by the participants during each of interrogation phases and then compute the percentage of true answers during that phase. Evaluating the participants' answers is done by listening to the experimental audio recordings. If the percentage of true answers is **strictly higher than 50%**, we consider that the participant was honest during that phase and deceptive otherwise.

Last, but not least, after annotating the data obtained from all participants, we had to annotate at least a part of the data using a second rater. We have chosen 3 random participants for cross-annotation, and the Cohen's kappa coefficient for inter-rater agreement had a value of $\kappa = \mathbf{0.921}$, corresponding to an almost perfect agreement between the two raters.

4.2.4 Participants

A number of 20 participants has undertaken this experiment, 7 of them being female and 13 being male, with an average age of 25.6 years. A majority of them (17 out of 20) had a background in Technical Sciences, while the three others had a background in Theatre, Physics, and Medical Sciences. We only had one main experimental group, as all participants were interrogated by a robot, so we did not have to divide the participants into various experimental subgroups. However, we analysed their personality profile as we wanted to make sure that no personality bias is present within our experimental group.

Personality profiling

For this experiment, we used the BIG5 personality test to evaluate the personality profile of our participants. This test allows us to evaluate 5 major traits: Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism. The test is comprised of 45 questions regarding the participants' own perception of themselves in various situations, with answers given on a 5-point Likert scale, with 1 corresponding to "Strongly disagree" and 5 to "Strongly agree". The test algorithm then computes a score from 1 to 5 for each of the 5 personality traits, with 1 corresponding to a low level and 5 to a high level. The analysis of the results over the 20 participants yielded the following results:

TABLE 4.1: BIG5 personality test analysis

Trait	Avg. ±std. dev.	Low (<3)	High (≥3)
Agreeableness	3.46 ±0.56	4	16
Conscientiousness	3.20 ±0.62	7	13
Extroversion	2.78 ±0.80	12	8
Openness	3.53 ±0.47	2	18
Neuroticism	2.96 ±0.97	9	11

Overall, we can notice that the average values for each personality trait are close to the average value of 3, the highest bias being for agreeableness, where the average score is of 3.53. On the other hand, if we separate the participants in subgroups depending on whether their score for each personality trait is higher or equal than 3 or strictly lower than 3, we notice that there is a very uneven distribution with respect to their agreeableness and openness, most of the participants being agreeable (16 out of 20) and open (18 out of 20). Therefore, we also added the separation into subgroups based on each personality trait as an experimental condition, in order to check if it had any impact on the participants' behaviour when lying.

4.3 Detecting deception using RGB-D cameras

RGB-D cameras, such as the Microsoft Kinect or Asus Xtion are nowadays ubiquitous in most robotic platforms. They allow the robot to acquire colour and depth information from its surroundings, including from their human interlocutor during HRI and, thanks to a series of algorithms and processing methods that will be detailed shortly, they also allow the robot to track a series of physiological and behavioural parameters that their human interlocutor exhibits. Throughout our work, we have three main research directions that involved using RGB-D cameras for lie detection:

- tracking the interaction distance
- tracking the eye position and openness in order to estimate the blink rate
- remotely estimate the heart rate using FFT analysis

Next, we will detail the techniques we used, the results we have obtained, as well as the limitations and issues we identified and the possible improvements that can be made to these techniques. Tracking the interaction distance, which also involves detecting the participants' faces as well as various facial features, is also necessary in order to be able to track the eye position and openness, as well as to remotely estimate the heart rate using FFT analysis. Therefore, we will detail this procedure first, also including the details that will be further relevant for the two other techniques that involve the use of RGB-D cameras.

4.3.1 Tracking the interaction distance

One of the first behavioural parameters we were interested in studying was the interaction distance between the two interlocutors (the human participant and the robot doing the interrogation). In particular, we were searching for a correlation between the interaction distance and the veracity of the participants' answers, using the mock-up crime and interrogation experimental setup detailed in the beginning of this chapter.

Measurement technique

In order to measure the distance in question, an Asus Xtion RGB-D camera is mounted on a fixed tripod in front of the SoftBank Pepper robot that is doing the interrogation. The camera is connected to the computer controlling the robot and overseeing the experiment and provides a live video stream. The RGB-D video stream is acquired using the **OpenNI2** ROS package, which splits the RGB and Depth streams and publishes them into two distinct ROS topics. Both streams are published at a VGA (640x480) resolution and at a frame rate of 30 Hz. From this point forward, the video streams are processed using our main software application, described at length in subsection 3.4.1, which subscribes to each of the ROS topics and runs a series of algorithms, that will be detailed shortly.

Firstly, we needed to use a state of the art, but open face tracker C++ library. We had a choice of several libraries, but we decided to use the CLM Face Tracker [60] , particularly due to our previous experience with it, as part of the research work done with Pauline Chevalier [88]. This library is able to track 66 facial landmark points in the RGB frame, corresponding to the coordinates various facial features (i.e., facial outline, eyebrows, eyes, nose, and mouth). A diagram of the tracked facial points and their indexing can be seen in the following figure (see Figure 4.3).

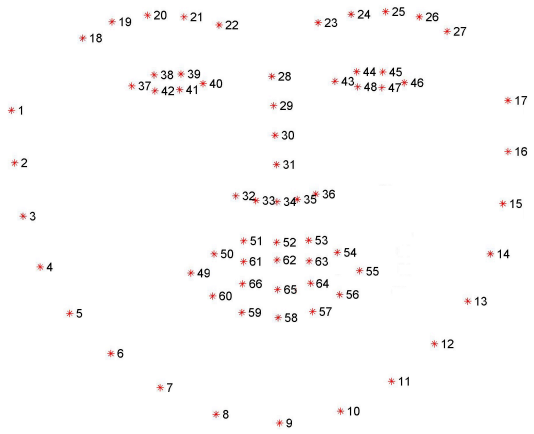

FIGURE 4.3: CLM Face Tracker landmarks

The library tracks in real time, frame by frame, the pixel coordinates of each of these 66 landmarks in the x-y coordinate system of the RGB video stream. It allows us to quickly extract the position and size (expressed in pixels and relative to the camera frame) of the following list of ROI:

- the bounding rectangle of the entire head
- the bounding rectangle of each eye
- the bounding circle of the inter-ocular area
- the bounding circle of the nasal area
- the circumscribed rectangle of the forehead area

For this part of the analysis, we are only interested in the bounding rectangle of the entire head, which is computed based on the minimum and maximum x and y coordinates of the entire list of facial landmarks. The position and size of this bounding rectangle allow us to extract the value that we are actually interested in:

the distance between the camera and the participant's head. In order to do that, we need to retrieve data from the other video stream provided by the Asus Xtion camera: **the depth stream**. This stream contains the exact distance between the camera and the point in space corresponding to a pixel that can be visualised in the RGB stream. Therefore, using the previously computed coordinates of the head bounding rectangle, we retrieve the distances corresponding to all the pixels inside the rectangle, and then compute the average distance over all these points. Last, but not least, the computed distance value is exported to a timestamped CSV file, that is analysed using the methods described in subsection 3.4.2.

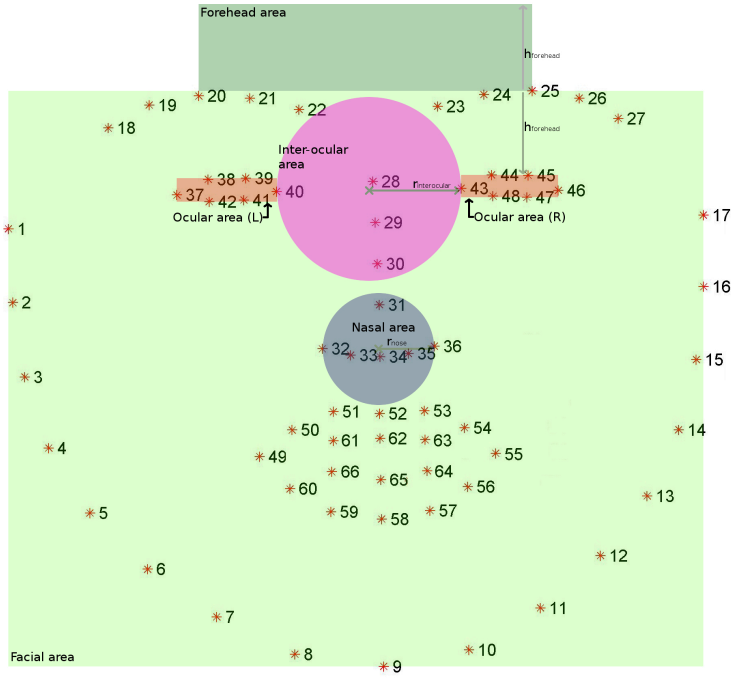

FIGURE 4.4: Facial ROIs and landmarks

Moreover, as tracking the face position and distance also allows us to implicitly track a series of facial landmarks corresponding to the eyes, eyebrows, nose and mouth, we can already compute the bounding rectangles and circles necessary for other parts of the analysis, as shown in Figure 4.4. Computing and storing the coordinates of the bounding boxes of each eye will allow us later to evaluate the eye position and openness. Then, the bounding box of the facial area as well as of the forehead area will be useful for attempting to detect the heart rate remotely, while the bounding circles of the inter-ocular and nasal areas will be later used in conjunction with the thermal data. The details of each procedure, as well as the facial landmark points used, will be detailed in the corresponding sections.

Results

Unfortunately, after processing the data from the 20 participants to the final version of our mock-up crime and interrogation experiment, we did not find any statistical correlation between the average value, standard deviation, or linear interpolation slope of the **interaction distance** and the veracity of the participants' answers, not even when taking into account their distribution into subgroups based on their personality profile.

This was in part due to the rather low number of participants and analysis samples (two for each participant, one for each analysis interval besides the reference interval) and in part to the hardware and software limitations. On one hand, the resolution of the Asus Xtion's RGB was low (640x480), which meant that the detected faces had at best a height of 100 pixels, and this had an impact on the tracking accuracy of the CLM Tracker library (and an even greater one when discussing the tracking accuracy for each of the facial features). Moreover, the Asus Xtion's RGB sensor had a high colour noise level, that was significant even in the best indoor lighting conditions that we were able to assure. The reduced detection and measurement accuracy were even more obvious in the first version of this experiment, where the analysis windows were much shorter and where detection noise had a stronger influence over the results, as the average, standard deviation and linear interpolation slopes were computed over a significantly lower number of samples.

4.3.2 Eye openness

Blinking is one of the physiological manifestations that have been proven to be correlated to deception in inter-human interactions, as discussed in chapter 2. Even though in most of the experiments previously conducted by psychologists, the timestamps of the blinks are manually annotated, we wanted to detect blinks automatically during our interaction scenarios. State of the art blink detection techniques, detailed in chapter 3, usually involve the use of video cameras, either placed in front of the person whose blink rate is being detected or on special glasses used to monitor ocular activity (e.g., gaze, pupil dilation, eye openness, and blinking).

As our research work relies heavily on the use of an RGB-D camera in conjunction with a face tracking algorithm that also detects a series of points corresponding to the contour of each pair of eyelids, it was in theory possible to detect individual blinks by monitoring the distance between the eyelids. However, due to the low camera resolution and high colour noise that the camera sensor had, we were not able to implement a robust and reliable blink detection algorithm, able to detect individual blinks and generate timestamps for each of them. Therefore, instead of computing accurate inter-blink intervals and blink rates, we decided to evaluate the participants' **eye openness** that would allow us to indirectly estimate the blink rate by evaluating its standard deviation.

Measurement technique

Once the coordinates of each facial feature point are being computed by the CLM Face Tracker algorithm that we have previously described, we process the coordinates of the ocular feature points. In particular, we are interested in tracking the y coordinate of the ocular feature points that correspond to the top and bottom of each eyelid:

- left eye: 38,39 (top) and 41,42 (bottom)
- right eye: 44,45 (top) and 47,48 (bottom)

We then estimate the average eye openness by subtracting the average y coordinate of the points corresponding to the top eyelids (38,39,44,45) from the average y coordinate of the points corresponding to the bottom eyelids (41,42,47,48). This value is exported to a timestamped CSV file, which is then analysed using the methods described in subsection 3.4.2. In particular, the standard deviation of the eye openness would be directly correlated to the estimated blink rate (a higher relative standard

deviation during a given interval meaning a higher estimated blink rate during that interval).

Results

After processing and analysing the data from the 20 participants during the final version of our mock-up crime and interrogation experiment, we did not find any statistical correlation between the average value, standard deviation, or linear interpolation slope of the **eye openness** and the veracity of the participants' answers, not even when taking into account their distribution into subgroups based on their personality profile. The reasons for which we assume we did not find any correlation are the same as in the case of the interaction distance, detailed in the previous subsection: the low camera resolution and the high colour noise, which meant that the measured inter-eyelid distances were of less than 10 pixels and that the tracking errors were of the same order of magnitude as the inter-eyelid distances. These were also the reasons for which we were unable to implement a robust blink detection in the first place and they will be later addressed by using a better camera and a better tracking algorithm.

4.3.3 Estimated heart rate

Heart rate (HR) is one of the major physiological manifestations that have been proven to be correlated to deception in inter-human interactions, even since the first attempts to detect deception using a polygraph approach. State of the art techniques of heart rate monitoring are divided between invasive, minimally invasive, and even non-invasive methods. Since in the beginning of our research we decided to focus purely on non-invasive measurement techniques, specifically the use of RGB-D and thermal cameras, we decided to also estimate the participants' HR remotely, using only these two cameras. After attempting to estimate HR using each of these two cameras, using a similar approach, we decided to use only the RGB-D camera, as it offered more robust results.

Measurement technique

Photoplethysmography (PPG) is a widely used heart rate measurement technique, detailed at length in chapter 3, that relies on the variation of the colour absorption properties of the skin as blood flows through it. Traditionally, in order to evaluate the colour absorption properties of the skin with a high accuracy, minimally invasive optical sensors are applied to the fingertips (e.g., in patient monitoring systems) or wrists (e.g., in smart watches and smart bands). Nonetheless, using the same base principle, we can estimate the heart rate remotely and by noninvasive means, using only a standard RGB camera.

As beat to beat detection using RGB video instead of a skin sensor is very hard to implement due to the low colour measurement accuracy and high colour noise, we had to extract the estimated heart rate using a post-experimental frequency domain analysis of the skin colour, in particular of its green component. The structure of the remote PPG detection technique we implemented is the following:

1. Identify a relevant region of facial skin.
2. Compute the value of the average green colour component for the skin area.
3. Log the raw computed value.

4. Extract the estimated heart rate post-experimentally using FFT analysis.

We start by identifying a region of facial skin, based on the previously extracted facial landmarks. We decided to identify two different face areas to compute the average green colour component of: one that uses the entire face as a reference and another that uses only the forehead area. The forehead area is smaller, containing less pixels, but is also less likely to contain pixels that do not correspond to the skin, while if using the entire facial bounding box as a reference, there will be a non-negligible number of non-skin pixels that will add a constant offset to the computation of the facial skin green component.

Computing the facial bounding box is straightforward and has already been detailed. In order to estimate the coordinates of the forehead, we use the coordinates of the centres of each eyebrow (points 20 and 25 in Figure 4.3 and Figure 4.4), as well as the coordinates of the ocular area (in particular, points 38, 39, 44, and 45). On the x axis, the forehead area is defined as the area between the x coordinates of the facial landmark number 20 and 25, while the bottom of the forehead area is computed as the minimum coordinate of the aforementioned pair of points associated to the eyebrows. The height of the forehead area is numerically equal to the vertical distance between the top of the eyebrows and the top of the eyelids.

Then, for each of these two ROIs, we compute the average green colour component and log it into a separate timestamped CSV file. These files are later processed post-experimentally using a frequency domain analysis procedure detailed in subsection 3.4.2. In particular, we will perform a FFT analysis of the facial green component and then filter the obtained spectrum in the 1...3 Hz frequency domain, that corresponds to potential heart rates of 60...180 BPM. The significant frequency peaks in that frequency domain are expected to correspond to the median heart rate of the participant during the measurement window. Moreover, the frequency resolution of the FFT analysis is inversely proportional to the window length:

$$f_{res} = \frac{f_s}{N_{samples}} = \frac{f_s}{f_s * length_{window}} = \frac{1}{length_{window}}$$

In the previous equation, f_{res} corresponds to the frequency resolution, f_s to the sampling frequency (which is of 30 Hz, same as the camera frame rate, after linearly interpolating the raw samples), $N_{samples}$ to the total number of acquired samples in the analysis interval and $length_{window}$ corresponds to the length of the analysis window expressed in seconds. Therefore, for a window length of 30 seconds we obtain a frequency resolution of $\frac{1}{30}$ Hz (2 BPM), while for a window length of 60 seconds we obtain a frequency resolution of $\frac{1}{60}$ Hz (1 BPM).

The main reason to modify the initial experimental setup

As we wanted to estimate the participants' heart rate with a resolution of at least 2 BPM (and ideally 1 BPM), we realised that we needed to ensure that the intervals on which we perform the FFT analysis had to be 30...60 seconds long. Therefore, we had no choice but to modify the initial version of the experimental setup, where each question-answer pair was analysed, having analysis intervals of around 10 seconds. Instead of this question-answer based analysis, we devised an interrogation procedure where questions are grouped into interrogation blocks and analysis is performed on each interrogation block. This allowed us to have interrogation windows that were longer than 60 seconds, allowing us to obtain the desired frequency resolution.

Results

After analysing the data obtained from the 20 participants to the mock-up crime and interrogation experiment, we obtained a series of correlations between the participants' relative estimated heart rate with respect to the reference interval and the annotated veracity of their answers for each interval, for three experimental subgroups: female participants, male participants, and participants with high neuroticism (over 3 out of 5 according to the Big5 Personality Test).

We first analysed if there were HR differences for the female participants while telling the truth and when lying. The data was normally distributed and was analysed with the ANOVA test. The median heart rate variability of the 7 female participants had a value of $-\mathbf{0.004}Hz(-\mathbf{0.24}BPM)$ when lying and of $+\mathbf{0.2}Hz(+\mathbf{12}BPM)$ when telling the truth, with $F(\mathbf{1,19}) = \mathbf{5.87}$ and $p = \mathbf{2.56} * \mathbf{10}^{-\mathbf{2}}$. This means the female participants had a higher heart rate when telling the truth compared to when lying.

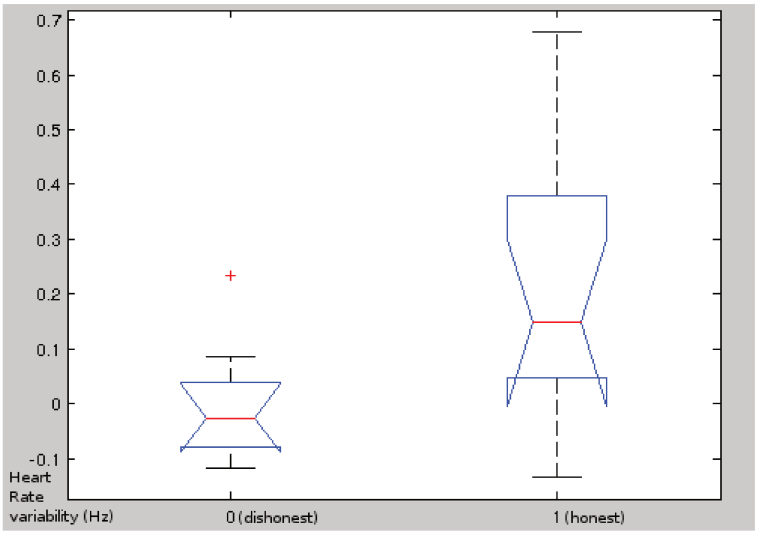

FIGURE 4.5: Female participants: heart rate variability

We also run the same test for male participants. The data was not normally distributed and was analysed with the Kruskal-Wallis test. The median heart rate variability of the 13 male participants had a value of $-\mathbf{0.185}Hz(-\mathbf{11.1}BPM)$ when lying and of $+\mathbf{0.01}Hz(+\mathbf{0.6}BPM)$ when telling the truth, with $\chi^{\mathbf{2}}(\mathbf{1,37}) = \mathbf{5.73}$ and $p = \mathbf{1.66} * \mathbf{10}^{-\mathbf{2}}$. This means that the male participants also had a higher heart rate when telling the truth than when lying.

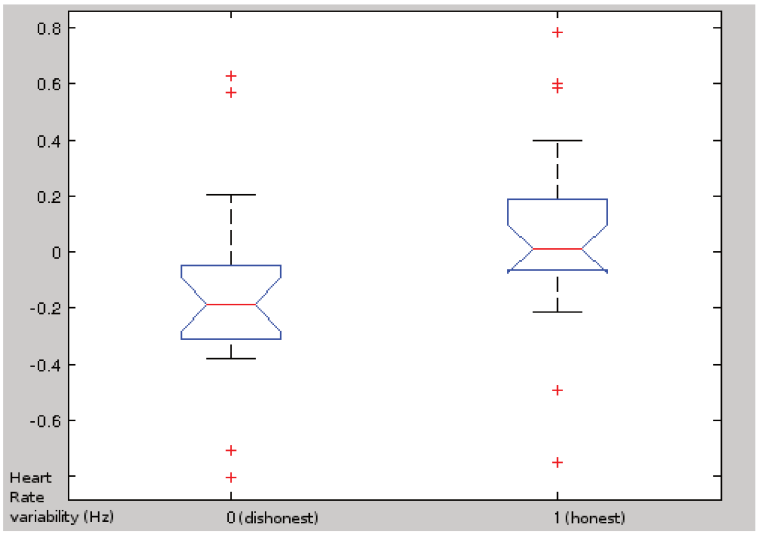

FIGURE 4.6: Male participants: heart rate variability

Furthermore, we analysed the HR differences for participants with high neuroticism when telling the truth and when lying. The data was normally distributed and analysed with the ANOVA test. The median heart rate variability of the 11 participants with high neuroticism had a value of $-0.021Hz(-1.26BPM)$ when lying and of $+0.167Hz(+10.02BPM)$ when telling the truth, with $F(1,31) = 5.28$ and $p = 2.80 * 10^{-2}$. This result means that the participants with a high neuroticism had a higher heart rate when telling the truth than when lying.

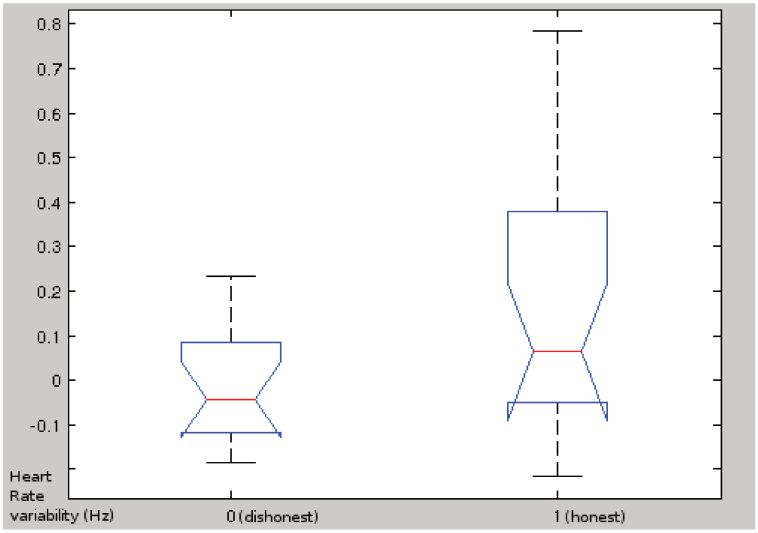

FIGURE 4.7: Participants with high neuroticism: heart rate variability

4.4 Thermal image analysis

State of the art deception detection methods in inter-human interactions also involve the use of thermal cameras. Firstly, they can be used to track the temperature variations in several facial ROIs, which are directly correlated to the changes in blood flow through those areas, that have been discussed at length in chapter 2. Secondly, they can be used to estimate the respiratory rate by tracking the area in front of the

nostrils, whose temperature varies as people breathe in and out, and then by using frequency domain analysis to extract the frequency of the periodic temperature variation. We therefore decided to track, by means of thermal imaging, the following physiological parameters:

- the temperature of the inter-ocular area
- the temperature of the nasal area
- the estimated respiratory rate (by means of FFT analysis of the nasal area temperature)

In this section, we will present our work related to detecting deception using thermal cameras, starting by detailing the procedure used to allow us to track the ROIs in the thermal camera frame, then continuing with the time domain analysis of the temperatures of the chosen ROIs and finishing with the frequency domain analysis that allowed us to estimate the participants' respiratory rates.

4.4.1 Tracking facial ROIs

One of the major challenges of measuring the temperature of specific facial ROIs is being able to track them accurately. If currently we can find a variety of free to use facial tracking libraries that use regular, RGB video, or that even manage to take into account the depth information from RGB-D cameras, there are no readily available facial tracking libraries that can detect facial features in thermal camera images. Therefore, we had two options if we wanted to autonomously track facial features in thermal imaging:

- manually train a generic object tracker (such as the one provided by the DLib library)
- track the facial landmarks in an RGB frame and then convert the coordinates to the thermal camera's frame of reference

Camera coordinate conversion

Since manually training a generic object tracker was time consuming and required large numbers of image samples, taking into account all the various head poses that could be observed during an in-the-wild HRI experiment, we decided to solve this problem by using the facial landmark points previously tracked in the RGB video stream and convert the RGB frame pixel coordinates to thermal frame pixel coordinates, using the camera calibration and coordinate converter package of the OpenCV library.

The main structure of our conversion algorithm is the following:

1. Compute the (X, Y, Z) real world coordinates of each facial landmark point observed in the RGB frame using:
    - the (x_{rgb}, y_{rgb}) RGB camera pixel coordinates of each facial landmark
    - the radial and tangential distortion coefficient vector (unique to every camera, useful to compensate for optical aberrations, obtained by calibrating the RGB camera)
    - the A_{rgb} camera matrix (a 3x3 matrix containing information about the focal length and the centre of the optical frame, obtained by calibrating the RGB camera)
    - the Z coordinate provided by the depth sensor of the RGB-D camera

2. Compute the (x_{th}, y_{th}) thermal camera pixel coordinates of each facial landmark point using:
    - the previously computed (X, Y, Z) real world coordinates of each facial landmark
    - the A_{th} camera matrix (same purpose as the one of the RGB camera)
    - the radial and tangential distortion coefficient vector (same purpose as the ones of the RGB camera)

The OpenCV library allows us to easily calibrate the camera matrices and the distortion coefficient vectors of RGB cameras, using a ready-to-use software package and procedure. This procedure involves the use of a black-and-white chequerboard (or another printed fixed pattern of known dimensions) that is moved around in front of the camera that needs calibration. Moreover, a similar procedure allows for the simultaneous calibration of two RGB cameras, allowing for an easy coordinate conversion from the frame of reference of one of the cameras to the other. However, as a thermal camera cannot detect light nor colours, and only temperature variations, this package and calibration procedure cannot be used as-is for calibrating both our cameras.

Workaround for thermal cameras

Therefore, we had to find a way to work around this limitation. We needed a calibration matrix that can be detected by both cameras. We realised that LED lights emit light, but also warm up while being lit. Therefore, we have built a LED matrix composed of 8x6 3 mm red LEDs, placed 70 mm apart, embedded in a black plastic panel, depicted in Figure 4.8.

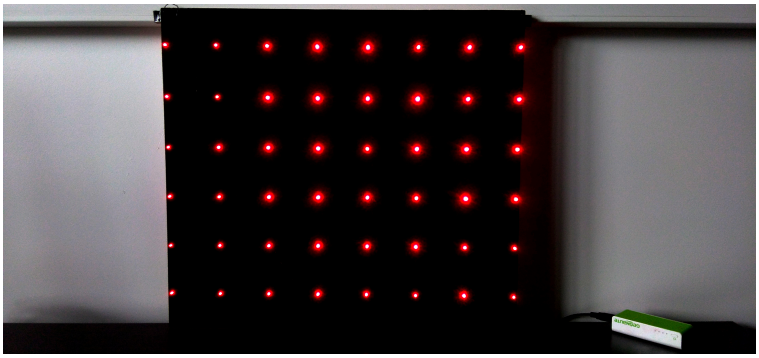

FIGURE 4.8: The LED matrix used for the camera calibration procedure

We have then created a software that detects the pixel coordinates of every LED in each of the two camera frames (RGB and thermal), based on the colour intensity and temperature differences with respect to the background, as shown in Figure 4.9:

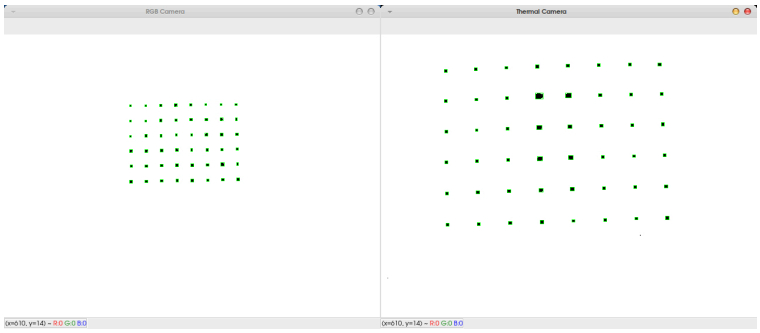

FIGURE 4.9: The interface of the LED matrix detection software

Then, after detecting the position of each LED in each of the two camera coordinate frames, we have used them as an input to the OpenCV camera calibration package, that later computed the conversion and calibration matrices we needed, as shown in the following figure:

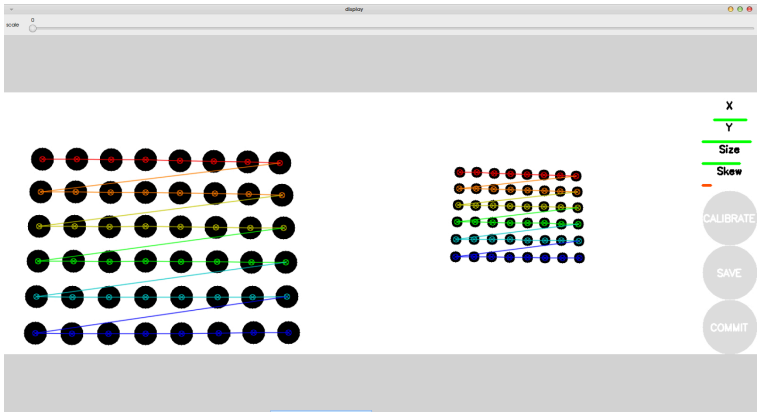

FIGURE 4.10: Using the OpenCV camera calibration package

Last, but not least, using the conversion and calibration matrices we needed, we implemented a coordinate conversion method in our main software application that tracks and logs the raw physiological parameters, allowing us to convert the pixel coordinates of any facial landmark detected in the RGB-D camera frame to pixel coordinates in the thermal camera frame. Moreover, we also created a series of methods that would quickly allow us to convert the coordinates of simple geometric shapes (such as rectangles or circles) from the RGB-D camera coordinate system to the thermal camera coordinate system. The following figure shows an example of the conversion done for the coordinates of all the facial landmark points, as well as of three circular ROIs (ocular and inter-ocular):

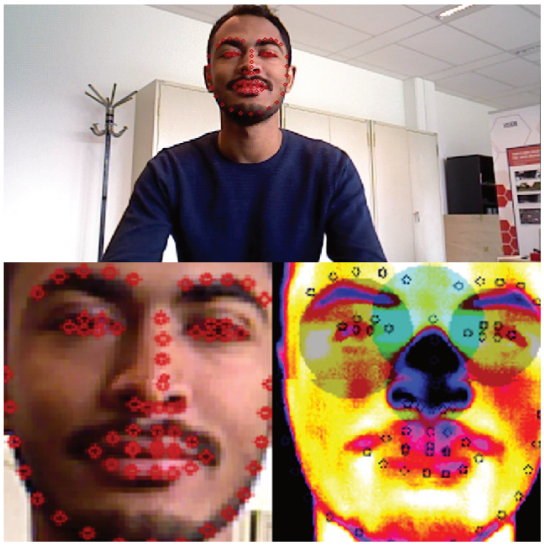

FIGURE 4.11: RGB and thermal face tracking - coordinate conversion

4.4.2 Facial area temperature variations

As we detailed in chapter 2, one of the physiological manifestations associated to deception is the variation in blood flow in several regions of the face, such as the nasal area and the inter-ocular area. An increased blood flow in these regions implies an increase in the skin temperature, relative to the rest of the face, that can be detected when using a thermal camera. We therefore decided to evaluate the relative temperature of these two regions throughout our experiment and search for correlations between the relative temperature variations and deception.

Measurement technique

Firstly, using the facial landmark points previously detected in the RGB frame, we identify the coordinates of the inter-ocular and nasal areas, as shown in Figure 4.4. In particular, for the nasal area, we used the coordinates of the points number 32 to 36, associated to the nose, while for the inter-ocular area we used the points corresponding to the inner eyelid corners (number 40 and 43).

In the case of the nasal area, we compute its centre based on the average x and y coordinate of the points number 32 to 36 (corresponding to the bottom of the nose), with its radius being half the distance between the extreme points (number 32 and 36). A similar approach was used to identify the inter-ocular area, computing the centre as the midpoint of the inner eyelid corners (points number 40 and 43) and its radius as half of the distance between these two points.

We then converted the pixel coordinates (centres and radii) of the two areas we have identified in the RGB frame into pixel coordinates in the thermal frame, continuing by computing the absolute average temperature of each of these regions. As we were interested in monitoring the evolution of the **relative** temperature of these regions with respect to the entire face, we also needed to compute the average temperature of the entire face. In order to do so, we converted the coordinates of the facial area (its bounding rectangle) in a similar fashion, then computed the average temperature of that region. Finally, we computed the relative average temperatures of the two aforementioned ROIs with respect to the face and logged these values to separate timestamped CSV files, that were analysed in the time domain.

Results

After processing and analysing the data from the 20 participants to our experiment, we did not find any statistical correlation between the average value, standard deviation or linear interpolation slope of the **relative temperature of the inter-ocular or nasal areas** and the veracity of the participants' answers, not even when taking into account their distribution into subgroups based on their personality profile. If we had issues with the accuracy of the previous measurements, that used only the RGB and depth sensors of our cameras, the tracking and coordinate conversion errors caused by the low camera resolutions, as well as by noise in both the RGB and depth streams, were amplified by the coordinate conversion procedure. Moreover, as the participants moved and especially turned their heads throughout the experiments, the projection of the tracked ROIs on the thermal image did not track exactly the same regions of skin, as the head angle with respect to the cameras changed.

4.4.3 Respiratory rate estimation

Much like the variations in heart rate, the variations in respiratory rate are one of the main physiological manifestations associated with deception. Moreover, even though the respiratory rate is traditionally monitored using invasive or minimally-invasive sensors, it can be also estimated remotely, using a thermal camera that monitors the flow of air in front of a person's nasal area, as they breathe in and out. In this subsection, we will detail our attempts to analyse the correlation between this physiological manifestation and deception in HRI.

Measurement technique

As a person breathes, they inhale and exhale periodically air at different temperatures. When breathing in, the air entering the nostrils is at room temperature, while the air that is being exhaled is at the body temperature. Facial skin, and especially the slightly moist tissue covering the nostrils usually has a temperature that is a few degrees lower than the body core temperature, but also significantly higher than our experimental room temperature (which is usually of around 20°C). Therefore, if we monitor the temperature of the nasal area, and especially the area situated in front of the nostrils, we will observe a series of periodic fluctuations of the temperature, with minimums when the person breathes in and maximums when the person breathes out.

We therefore performed a frequency domain analysis on the previously computed and logged relative temperature of the nasal area with respect to the average face temperature, using the same principle used to estimate the participants' HR using an FFT analysis of the average facial green colour component. In this case, we have chosen the 0.1...1.0 Hz frequency interval for filtering the resulting spectrum of the FFT analysis, as we considered that it would correspond to a normal breathing rate for a person sitting down.

Results

As it was the case for the analysis of the facial area temperature variations, no correlations between the participants' estimated respiratory rate and the veracity of their answers were found. We believe the lack of correlations was due to the poor measurement accuracy of our system, mainly generated by the low camera resolution and camera noise.

4.5 Post-experimental questionnaires

4.5.1 Objectives and questions

Our experimental design included a post-experimental questionnaire, comprised of questions concerning the extent to which they found the experimental design appropriate, the feelings and emotions they resented during the experiment, as well as their perception of how difficult they found lying to their interlocutor. We are particularly interested in evaluating the following aspects:

- the participants' perception of their stress level during the experiment
- their desire to win the assigned reward
- their perception of their interlocutor's behaviour and abilities
- their performance throughout the game

As previously shown in chapter 2, some of the physiological manifestations associated to deception can also easily be associated to stress and frustration. It is therefore important to minimise the magnitude of these emotional responses, as well as to understand to what extent the participants feel they have exhibited them during the interaction. On top of that, we also want to ensure that the participants feel motivated to attain their objective, therefore to lie, to an extent that can be compared to a real-life scenario, where lying would bring them a reward they actually desire. Of course, in order to evaluate whether we succeeded or not in motivating them, we have to evaluate their perception and motivation.

For the final version of our experimental setup, we have devised a questionnaire containing 11 questions, as well as basic information about the participant (full name, age, country of origin, sex, career field and whether they had already taken part in an interaction with a robot). The question list they had to answer on a 7-point Likert scale, with 1 corresponding to "Strongly disagree" and 7 to "Strongly agree", is the following:

1. Did you want to keep the chocolate bar?
2. Do you think the interrogator knew what you did inside the first room?
3. Do you think you gave up to the interrogator any information that you were not allowed to?
4. Do you think the interrogator was able to detect when you were telling the truth and when you were not?
5. Did you find the first part of the experiment stressful?
6. Did you find the interrogation stressful?
7. Do you think a human interrogator would have been more able to detect when you are telling the truth and when you are not?
8. Do you think it would have been easier for you to lie to a human interrogator?
9. Were you afraid you were going to lose the reward in the end?
10. Did you feel threatened by the interrogator at any moment?
11. Do you think the questions you were asked were important in order to establish what you did in the first experiment room?

4.5.2 Analysis of results

The 20 participants that took part in the final version of our mock-up crime and interrogation experiment also filled in the post-experimental questionnaires. For each of the question, we computed the average value and standard deviation of their answers, with 1 being the minimum value and 7 the maximum value, obtaining the following results:

TABLE 4.2: The results of the mock-up crime & interrogation post-experimental questionnaire

Question	Avg. ±std. dev.
1. Did you want to keep the chocolate bar?	5.375 ±1.907
2. Do you think the interrogator knew everything that you did inside the first room?	5.125 ±2.071
3. Do you think you gave up to the interrogator any information that you were not allowed to?	3.917 ±2.283
4. Do you think the interrogator was able to detect when you were telling the truth and when you were not?	3.708 ±1.732
5. Did you find the first part of the experiment stressful?	1.667 ±1.204
6. Did you find the interrogation stressful?	3.792 ±1.933
7. Do you think a human interrogator would have been more able to detect when you are telling the truth and when you aren't?	4.500 ±1.794
8. Do you think it would have been easier for you to lie to a human interrogator?	2.625 ±1.245
9. Were you afraid you were going to lose the reward in the end?	3.458 ±1.956
10. Did you feel threatened by the interrogator at any moment?	2.958 ±1.546
11. Do you think the questions you were asked were important in order to establish what you did in the experiment room?	4.875 ±1.676

Overall, we concluded that the participants found the reward desirable (5.375 ±1.907), even though we felt this aspect could have been improved for our next experimental design. Moreover, in spite of having concealed the hidden camera from sight, the participants were also rather convinced that the interrogator already knew what they did during the first part of the experiment (5.125 ±2.071). We did our best to make them believe there was no way anyone could have known what objects they had hidden before the experiment was over, but they were rather sure this was not the case nonetheless. However, they had a significantly less strong belief that the robot interrogator was able to detect when they were lying and when they were telling the truth, considering its abilities rather average (3.708 ±1.732). In line with this belief, they also considered a human would have been slightly more able to detect deception (4.500 ±1.794) and also that it would have been far less easier to them to lie to a human interrogator (2.625 ±1.245). This can be easily justified by the fact that they simply did not expect the robot to already have lie detection capabilities.

When analysing their emotional response throughout the experiment, we noticed that even though they considered the mock-up crime not stressful at all (1.667 ±1.204), their stress level during the interrogation phase was slightly under average (3.792 ±1.933). We tried to ensure that their perceived stress level is as low as possible during this experiment, in spite of expecting them to lie, and this result showed that we managed to accomplish this objective, even though the perceived stress level could have been even lower.

4.6 Summary

In this chapter, we have discussed our attempts to detect deception using solely non-invasive sensors: an RGB-D and a thermal camera. We have detailed the experimental methodology, the data measurement, and analysis procedures used to collect the raw physiological measurements, as well as the results of our post-experimental analysis. In particular, we analysed the participants' interaction distance with respect to the robot undertaking the interrogation, their estimated heart rate, their estimated blink rate, as well as the relative temperature of their inter-ocular and nasal areas and their estimated heart rates. We attempted to find correlations between the

evolution of these parameters during the second part of our experiment (the interrogation procedure) and the veracity of their answers during each of the interrogation phases.

4.6.1 Results

We have found a series correlations between the participants' estimated heart rate and the veracity of their answers during the interrogation. In particular, for three different experimental subgroups (female participants, male participants, and participants with high neuroticism), the participants' estimated heart rate was higher when lying than when telling the truth. These results have been published and presented at the RO-MAN 2018 conference (27-31 August 2018, Nanjing, China), in the article entitled *First attempts in deception detection in HRI by using thermal and RGB-D cameras*.

Moreover, the analysis of the feedback received from our participants shows that our experimental design managed to accomplish its objectives, as the participants found they were enticed to obtain their reward and as their stress level was very low during the first part of the experiment and also below average during the interrogation phase of the experiment.

4.6.2 Limitations and future developments

Detecting deception using only non-invasive measurement devices and methods is a challenging task, especially in experimental setups that can be considered in-the-wild. On one hand, the limited resolution and accuracy of video cameras (whether they measure colour, depth, or temperature) implies that the precision with which we can measure any physiological manifestation is also limited. Moreover, our attempt to make in-the-wild experiments, where the participants' movements, reactions, or behaviours are not forced or hindered in any way by the experimenter, also means the participants may move in ways that corrupt the data acquisition process (for examples, head poses during which the facial ROIs we track no longer have the same 2D projection as in the case of a full frontal pose). Last, but not least, frequency domain analysis by means of FFT forces us to have large analysis windows, during which the parameter we monitor might have significant variations that we cannot monitor.

Some of these issues could be solved by reducing the amount of freedom of participants throughout the experiment or by asking them to follow strict rules when it comes to their behaviour. For example, participants could be asked to always look at the cameras, to never cover their faces with their hands or to limit their movements; we could also slow down the dialogue between the interrogator and the participant, while adding large pauses that would allow us to acquire more data during the interaction. Nonetheless, we consider that these approaches would be a breach of our experimental guidelines and that, as a consequence, would induce a drastic bias in our research.

Therefore, at this point we decided to improve the precision of our measurement techniques and devices. On one hand, we decided to use a better quality and higher resolution RGB camera instead of the Asus Xtion RGB-D camera, coupled with a more robust face tracking algorithm. On the other hand, we decided to explore the possibility of using minimally-invasive sensors and devices, in the form of smart watches or arm/wrist bands, that would allow us to monitor some of the participants' physiological manifestations with a far better precision and accuracy, while

not hindering in any way with their ability to interact normally or perform regular tasks. Last, but not least, we decided to develop an experimental scenario where deception is elicited in a less formal context than that of a mock-up crime and interrogation. In the following chapter, we will detail our attempts to improve our ability to detect deception, as well as our results.

4.7 My contribution

The main contributions to this chapter were in the design and implementation of the experimental setup, the software used to collect and analyse the physiological parameters studied throughout the experiment, the LED matrix and additional software used to calibrate the thermal - RGB cameras coordinate conversion system, as well as the post-experimental questionnaire used to evaluate the participants' experience during the experiment.

Chapter 5

Using wearable sensors in conjunction with an RGB camera

5.1 Introduction

In the previous chapter, we presented our attempts to detect deception using only noninvasive measurement devices and techniques, in particular an RGB-D and a thermal camera. As it was concluded at the end of chapter 4, the lack of more comprehensive results was partly due to the limitations of the hardware and software used for noninvasive detection. Therefore, for this part of our research, we decided on one hand to use improved noninvasive detection devices and algorithms, particularly by replacing the previously used RGB-D camera with a high-definition RGB camera, as well as to complement the existing noninvasive techniques with the use of a **minimally-invasive wireless armband**.

Moreover, we improved our experimental setup by attempting to make it less stressful and less formal, while still following the experimental design guidelines we have established in section 3.2. Hence, our new experimental setup takes the form of a **card guessing game**, detailed at the beginning of this chapter, played twice by each participant, once with a human and once with a robot game partner. Then, the results concerning the relationships and correlations between deception and the participants' **head position and orientation**, **eye openness**, **heart rate** and **skin conductance**, using the new and improved sensors and detection techniques will be presented. The results of the post-experimental and post-round questionnaires that we used to evaluate the participants' perception of the experiment will be detailed. The chapter ends by a summary of the results, challenges, and limitations of this work.

The experimental setup and detection techniques detailed in this chapter are also referred to in chapter 6, where we discuss the differences between the interactions with a human and a robot game partner.

5.2 Experimental setup: the card guessing game scenario

Our second experimental setup was designed using a different approach. Instead of a mock-up crime and interrogation approach, we wanted to create a less stressful and more entertaining interaction scenario, but still one during which the participants were motivated to lie at given moments. Therefore, we designed a card game that, on one hand, had rules that were easy to understand and learn by the participants and on the other hand can also be easily played by a robot game partner. Moreover, since the study of the physiological and behavioural differences between the deception of a human and of a robot interlocutor was an important part of our

research that we had not explored yet, we designed two game conditions: one with a robot game partner and the other one with a human one. These differences are discussed at length later in chapter 6.

5.2.1 The decks of cards

Two decks of cards were used to play this game. The first one comprised 32 square cards, with light blue back sides, each of them displaying a picture, as well as French and English subtitles with the name of the item represented in the picture. The 32 pictures contained the following various objects, animals, and famous persons:

TABLE 5.1: The 32 items displayed on the picture cards

plane	smartphone	Albert Einstein	shark
helicopter	TV	Elvis Presley	crocodile
X-Wing Starfighter	old radio	Marylin Monroe	tiger
The Millenium Falcon	old computer	Amy Winehouse	wolf
car	hammer	Donald Trump	cow
bicycle	screwdriver	Zinedine Zidane	horse
Batmobile	vase	Scarlett Johansson	zebra
AT-AT Transport	bronze statue	Serena Williams	deer

FIGURE 5.1: The two decks of cards

The second deck of cards contained 8 smaller cards, with light green back sides. Half of them had a red front side and instructed the player to cheat on their game partner, while the other four cards had a green front side and instructed the player to help their game partner. This deck of cards was shuffled before each game was played and then placed face down on the table.

5.2.2 The game rules

Each of the 8 game rounds, the participant started by choosing one of the 32 image cards, and then proceeded by drawing the top card from the small deck on their right. Next, after the participant looked at the small card they had drawn and confirmed they were ready to continue, the game partner started asking a series of 5 questions (according to a question tree can be found in Appendix C) attempting to guess what was the item displayed on the chosen image card. The participant had to answer each of these questions with *Yes* or *No*, but their objective depended on the colour and instructions of the small card they had drawn. If the participant had drawn a green small card, they were supposed to help their game partner guess what

was shown on their chosen image card, therefore answering the questions truthfully. Otherwise, if the participant had drawn a red small card, they were supposed to mislead their game partner into making a wrong guess.

After these 5 questions, the robot or human game partner attempted to guess what was represented on the chosen card and the player confirmed whether the guess was right or wrong. Then, the player was asked to show their game partner the two cards and, if they succeeded in accomplishing their round objective, they won a point. More specifically, if they had drawn a small green card and their game partner ended up guessing the correct image, the player won a point, as well as if they had drawn a small red card and their game partner ended up making a wrong guess.

5.2.3 The experiment

Each participant was instructed by the experimenter what the game consisted of, as well as the potential rewards for playing it well. More specifically, if at the end of the game the player's score was of 4 points or more out of 8, they were given marshmallows or candies, while if they achieved the maximum score of 8 points, they entered a lottery for a 50€ gift card, held with all the other participants who also scored the maximum score. Afterwards, they were required to sign the consent form, and then were asked to sit down at the game table, in front of their game partner, which was either a SoftBank Pepper robot or a human. A small box, useful for discarding the used cards, a sheet of paper with the game rules written in English and French, as well as a post-round questionnaire were also placed on the table, between the two decks of cards.

As the experiment started, the participant's game partner explained them in details the rules of the game, and then they were asked if they were ready to start playing the game. Once the participant agreed, the first round of the game begun, with the game partner giving verbal cues at each stage of the game and telling the participant what they were supposed to do next. After the players were notified whether they won a point or not, they were also told what their current score was, as well as how many rounds were played. Moreover, at the end of each round, the participant was asked to fill in a short post-round questionnaire comprised of three questions concerning the round that had just ended.

FIGURE 5.2: Playing the card game with a robot game partner

At the end of the 8 game rounds, the participants were told by their game partner what their final score was, as well as what they ended up winning. Those who managed to achieve the maximum score were told they had entered the lottery for the gift card and that they will be notified by e-mail later on whether they were the winners or not. Lastly, they were notified that the experiment was over and then asked to fill in a post-experimental questionnaire.

It is important to mention at this point that, in spite of the human experimenter and the robot having very different heights when standing up (about 180 cm for the human experimenter and 120 cm for the Pepper robot), we ensured that the heads of the two game partners were placed at the same position during the experiment. The human game partner was therefore sitting down, while the robot was standing up during the game, with their two heads being at the same height. Moreover, the camera was placed at the same spot relative to the table in both cases, while the human experimenter was also sitting down at the same spot where the robot was previously standing up. This was done to eliminate any source of bias with respect to the participants' head position and orientation throughout the games that would be due to the robot's or human's position in front of them.

5.2.4 Annotating the experimental data

The data annotation process was very similar to the one we had used in the previous experiment, using the audio recordings of the experiments. First of all, we had identified the beginning and the end timestamps of each relevant experimental interval, starting with the **baseline interval**. The baseline interval begun when the game partner started explaining the game instructions to the participant after the latter had sat down at the table. Then, we identified 8 analysis intervals, one for each round of the game. Each game interval started when the game partner started asking the first question to the player and ended after the player showed the game partner the two cards chosen for the round.

The other purpose of the annotation process was the evaluation of the truth value of the participants' statements. In this experimental scenario was designed to evaluate if the main purpose of the participant during a round was to deceive or to be honest with their game partner. The procedure was therefore simple and did not require us to analyse the veracity of every single answer. If the participants had drawn a green card and ended up scoring a point, they were considered honest during that round, while if they had drawn a red card and had also ended up scoring a point, they were considered deceitful during that round, regardless of how many questions they answered truthfully. In the rare occurrence of a participant not scoring a point during a round, the data was discarded and not used for the statistical analysis, as it was considered inconclusive.

All the experimental data was first annotated by the experimenter, yet we also had to annotate a part of the data using a second rater. We had chosen 4 random participants for cross-annotation, and the Cohen's kappa coefficient for inter-rater agreement had a value of $\kappa = \mathbf{0.938}$, corresponding to an almost perfect agreement between the two raters.

5.2.5 Participants

A number of 25 participants had undertaken this experiment, 4 of them being female and 21 being male, all of them having a background in Technical Sciences. Each of them played the game twice, once with a robot game partner and once with a human

game partner. They were distributed into two experimental groups: 13 of them played the first game with a robot partner, while the other 12 played their first game with a human partner. The 4 female participants were evenly split between the two experimental groups, to reduce to a minimum any bias induced by the participants' gender. Moreover, we also had to take into account the participants' personality profile when distributing them into experimental groups.

Personality profiling

For this experiment, we decided to replace the previously used BIG5 personality test with the short, revised version of the Eysenck Personality Questionnaire (EPQ). We asked the 25 participants to fill in this questionnaire as part of our card guessing game experiment and we established their EPQ-L dimension (predilection to lie), on a scale from 0 to 12, based on their answers. We analysed the average and standard deviation of the EPQ-L for each of the two experimental groups, defined by who they played the first game with, and we obtained the following results:

TABLE 5.2: Eysenck personality questionnaire analysis

Group	Avg. ±std. dev.	Low ($\leq$6)	High (>6)
Human player first	6.42 ±1.88	8	4
Robot player first	6.15 ±2.23	6	7
Overall	6.28 ±2.03	14	11

As it can be observed, the average values of the participants in each of the two experimental groups were rather close to one another (H = 6.42, R = 6.15) and also close to the middle of the measurement scale (6 out of 12). However, if we count the number of participants with a low predilection to lie (less or equal to 6 out of 12) and those with a high predilection to lie (higher than 6 out of 12), we can notice that there is an unbalance in the group that played the game firstly with a human player. Nonetheless, the same way we used the BIG5 personality traits as experimental conditions in the post-experimental analysis procedure, we used the membership to these two subgroups (low or high predilection to lie) as an experimental condition for the post-experimental analysis procedure.

5.3 Detecting deception using RGB cameras

In the previous chapter, our attempts to track several physiological parameters using an Asus Xtion RGB-D camera were discussed. Two of these parameters, the eye openness and the estimated heart rate, were measured using only the RGB sensor of the RGB-D camera, while the interaction distance was the only one monitored using the depth data stream provided by the camera. Since our previous results left a lot of room for improvement, we decided to use a higher resolution and higher quality RGB camera sensor instead of the previously used RGB-D camera, while trying to monitor even more parameters than before using this new camera.

In particular, we decided to monitor the following set of parameters and to search for correlations between them and the participants' honesty during the experiment:

- head position and orientation
  - x,y position *(camera frame pixel coordinates)*
  - height and width *(camera frame pixel coordinates - correlated to the interaction distance)*

  - roll, pitch, and yaw

- eye openness

Compared to the previous experiment, we extended our analysis of the head position and orientation, using only the RGB data provided by our Logitech C920 camera, continuing to also track the eye openness in order to estimate the participants' blink rate. In the following pages of this section, the techniques we used to measure the aforementioned physiological parameters are detailed, as well as the results obtained when searching for correlations between their evolution and the veracity of the participants' answers during each of the round games.

5.3.1 Tracking the head position and orientation

If in the first experiment we only focused on the analysis of the participants' interaction distance (the distance with respect to the camera placed in front of their robot interlocutor), for this experiment we decided to extend this analysis to several parameters. More specifically, we tracked, in the pixel coordinate system of the RGB camera, the participants' head x and y coordinates, and their head height and width, which is inversely proportional to the interaction distance. Moreover, we computed their head roll angle and estimated their head pitch and yaw angles based on various facial feature landmark points. We also attempted to correlate the evolution of these parameters during each of the 8 round games, relative to the baseline interval.

Measurement technique

The Logitech C920 camera used to monitor the participants' head position and orientation was placed on a fixed bracket in front of the game partner (either the SoftBank Pepper robot or the human experimenter). The camera was connected to the computer controlling the robot and overseeing the experiment, providing a live video stream acquired using the **usb_cam** ROS package. The stream was published at a HD (1280x720) resolution and at a frame rate of 30 Hz. This resolution was preferred to the Full HD resolution (1920x1080) that the camera was capable of due to the significant increase in the computation power needed to process the 1920x1080 frames for data analysis. This stream was further processed by the main software application we described in subsection 3.4.1, which subscribed to the camera's ROS topic and run the algorithms necessary to analyse the video footage. These algorithms are presented shortly.

We used a very similar approach to the one used in the previous chapter. We used an open-source state of the art face tracking library, the DLib Face Tracker. This replaced the previously used CLM Face Tracker, as it has improved tracking performance and robustness. The library provides a very similar output to the one provided by the CLM Face Tracker library, tracking 68 facial landmark points that correspond to various facial features (facial outline, eyes, eyebrows, nose, mouth), as detailed in the following Figure 5.3:

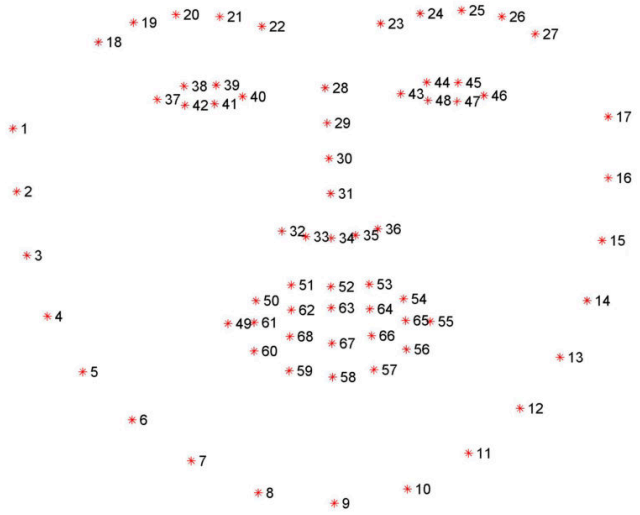

FIGURE 5.3: DLib Face Tracker landmark points

Most of the points' indices correspond to those previously used, the extra two points corresponding to the corners of the lips. The library tracks the x and y pixel coordinates of each of these 68 landmarks in the camera's pixel coordinate system. We used the coordinates of these points to compute or estimate the head position and orientation coordinates we were interested in.

The head position and size were computed in a similar fashion as previously. At each frame, the minimum x and y coordinates were computed over the 68 facial landmarks, therefore defining the bounding rectangle for the entire face. The head width and height correspond to the difference between the maximum and minimum x and y coordinates, respectively, while the head centre coordinates were computed as the average of the minimum and maximum x and y coordinates previously computed. The face distance wes not computed anymore, since the RGB camera did not provide depth information, but the head width and height were both inversely proportional to the interaction distance. Therefore, any significant variations of the measured head height and width during the experiment implied significant variations of the estimated interaction distance. All these parameters (head x and y coordinates, head width, head height) were measured in pixels, with the y axis being inverted due to the origin of the camera reference system being placed in the top-left corner of the image.

The head orientation was estimated based on 3 angles: the **roll**, the **pitch**, and the **yaw**. Since the RGB camera only provides a mostly frontal 2D projection of the participants' face, we can only compute the head roll accurately, while the pitch and yaw can only be estimated. Again, since our interest resided in determining the variations of these angles throughout the experiment, we can safely use an estimation of these angles instead of accurately computing their values. For all these three angles, the reference facial landmarks were those of the eyes (points 37-42 for the left eye, points 43-48 for the right eye) and those of the nose (28-31 for the nasal bridge, 32-36 for the tip of the nose). Based on these reference points, we computed the x and y coordinates of the centre of each eye by averaging the x and y coordinates of each of the landmark points that corresponded to them, as well as the x and y coordinates of the centre of the nose tip.

The head roll was computed based on the angle of two lines: the line connecting the two centres of the eyes, as well as the line interpolating the 4 nose bridge

landmark points. In the case of a symmetric face whose frontal plane was also perfectly aligned with the camera, these two lines were perpendicular to one another. However, since the head pose varied throughout the experiment, we decided to add a redundancy and computed the head roll as the average between the angle of the nasal bridge line and the perpendicular to the inter-ocular line. This angle was measured in radians, with respect to the vertical axis of the camera frame.

Then, the face yaw was expressed as the distance in pixels on the x (horizontal) axis between the midpoint of the two eyes (whose centre coordinates was previously computed) and the centre of the nose tip. Similarly, the face pitch was expressed as the distance in pixels on the y (vertical) axis between the midpoint of the two eyes and the centre of the nose tip. These values were expressed in pixels and their evolution was correlated to the evolution of the real face pitch and yaw angles.

The position and angle coordinates were logged into two individual CSV files, one for the position coordinates and the other for the angle coordinates. Their average values, standard deviations, and slopes of the linear interpolation over each interval relative to the reference interval were analysed post-experimentally using the methods described in subsection 3.4.2.

Results

After analysing the data obtained from both games played by the 25 participants to our card guessing game experiment, we obtained a series of correlations between several parameters defining the participants' head pose and the veracity of their answers during each round. In particular, we found the head height (and therefore the interaction distance), the head vertical position as well as the head roll angle to be correlated to the honesty of the participants throughout the rounds. Some of the results had been observed only in one of the experimental conditions and the entire list of results can be summed up in the following table, with the median value of each parameter during the rounds when the participants were being deceptive being compared to its median value during the rounds when they are honest:

TABLE 5.3: Head position and orientation: overall results

Parameter	Measure	Value when lying	EPQ-L	Partner	Test
head height	slope	higher	all	both	t-test
head height	slope	higher	all	human	t-test
head height	slope	higher	high (>6)	both	ANOVA
head y coord.	average	higher *(inv. y axis)*	all	robot	ANOVA
head roll	slope	lower	low (≤6)	both	t-test

The relative **linear interpolation slope** of the participants' head height had a median value of $+\mathbf{5.12 * 10^{-2}}$ when they deceived and of $+\mathbf{2.67 * 10^{-2}}$ when they were being honest with their game partner, with $\boldsymbol{p = 3.73 * 10^{-3}}$, for an analysis done on **all the participants** and with **both game partners**. The data was normally distributed and analysed with the t-test. The measured head height was inversely proportional to the interaction distance, so this result implies that the participants approached their heads to the game partner towards the end of the question series and did so at **a higher rate** when deceiving their game partner compared to when they were honest with them. The same parameter was also analysed and yielded

correlations on some of the experimental subgroups: the games played with a human game partner, as well as both games played by the participants with a **high EPQ-L** value (>6 out of 12), as shown in Table 5.4:

TABLE 5.4: Head height relative linear interpolation slope: subgroups analysis

Partner	EPQ-L	Test	$\tilde{h}$: lie	$\tilde{h}$: truth	p-value
human	all	t-test	$\mathbf{7.26 * 10^{-2}}$	$\mathbf{1.26 * 10^{-2}}$	0.001
both	H (>6)	ANOVA	$\mathbf{6.86 * 10^{-2}}$	$\mathbf{1.81 * 10^{-2}}$	0.046

Both data sets were normally distributed, tested with t-test and ANOVA, respectively, and the results of the analysis confirmed the previous conclusion that we had drawn on the entire data set. The cells coloured in light green indicate which of the two median values (for truthful or deceptive rounds) was higher.

The the participants' relative **average face y coordinate** had a median value of was $\mathbf{-7.75 * 10^{-1}}$ pixels when they deceived and of $\mathbf{+3.78 * 10^{-1}}$ pixels when they were being honest with their game partner, with $\boldsymbol{F(1, 187) = 4.04}$ and $\boldsymbol{p = 4.59 * 10^{-2}}$, for an analysis done on all the participants, but only on the games played with a **robot** game partner. The data was normally distributed and was analysed with the ANOVA test. As the y axis was inverted (i.e. lower y coordinate values mean a higher vertical position of the object), the results imply that the participants lifted their heads higher when lying than when telling the truth.

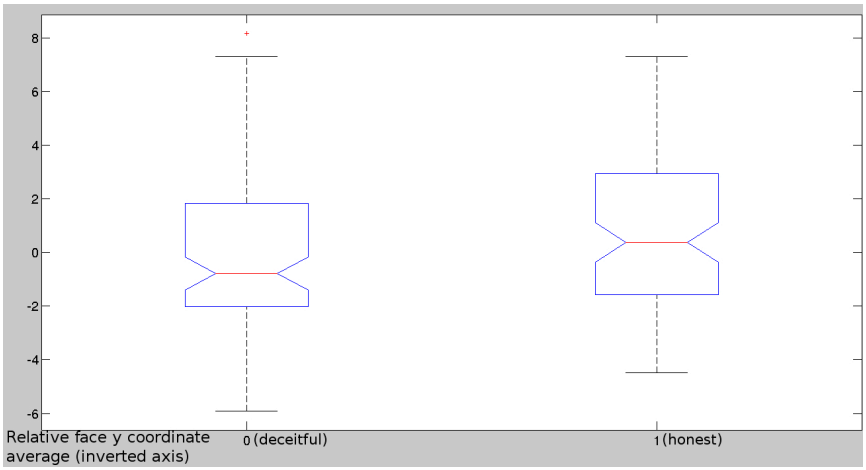

FIGURE 5.4: Relative face y coordinate average (robot game partner)

The relative **linear interpolation slope** of the participants' **head roll** had a median value of $\mathbf{-1.61 * 10^{-3}}$ when they deceived and of $\mathbf{-2.38 * 10^{-3}}$ when they were being honest with the game partner, with $\boldsymbol{p = 4.10 * 10^{-3}}$, for an analysis done only on participants with a **low EPQ-L** value ($\leq$ 6 out of 12) and both game partners. The data was normally distributed and was analysed with the t-test. Both values were negative and, due to the y-axis inversion, the results imply that the head roll angle **increases in a clockwise direction** towards the end of the card guess questions sequence. Moreover, the angle increases so at a higher rate when participants are being honest than when they are deceiving their game partner.

According to several studies [47] [48] [49], an over-controlled body posture has been previously associated with a deceptive behaviour. Our results (the more upright head position, the lower increase of the head roll angle and the higher decrease of the interaction distance) when the participants are being deceptive, are also signs of a more controlled body position, so these conclusions are in line with the literature.

5.3.2 Eye openness

In our previous experiment, we had unsuccessfully attempted to track the participants' blink rate. However, since we wanted to somewhat estimate this parameter and correlate it with the veracity of the participants' statements, we developed a method to track the eye openness instead of detecting individual blinks. In the case of the research and the setup presented in the current chapter, we decided to evaluate again the participants' eye openness and to attempt to detect individual blinks. Nonetheless, in spite of a higher camera resolution and of a more robust face tracking library, we still did not manage to detect individual blinks with a satisfactory accuracy. We have developed an algorithm that used a Finite State Machine approach, based on the measured distance in pixels between the facial landmark points corresponding to the top and bottom eyelids, and detecting when the distance between the eyelids was over or under a dynamically-adjusting threshold. Regardless of how we tuned our algorithm, it was still detecting a significant proportion of false positive or false negative samples. Therefore, we focused simply on evaluating the participants' eye openness, as we did in the previous experiment.

Measurement technique

The measurement technique is identical to that discussed in subsection 4.3.2 of the previous chapter. In this setup, we use the following ocular landmark points provided by the DLib Face Tracker library, which have the same indices as before, to compute the distance between the eyelids:

- left eye: 38,39 (top) and 41,42 (bottom)
- right eye: 44,45 (top) and 47,48 (bottom)

The distance between the eyelids is computed by subtracting the average y coordinate of the points that correspond to the top of the eyelids (38, 39, 44 and 45) from the average y coordinate of the points that correspond to the bottom of the eyelids (41, 42, 47 and 48). This raw value is then exported to a timestamped CSV file, which is then analysed using the methods described in subsection 3.4.2. We study in particularly the standard deviation of the eye openness, which is directly correlated to the estimated blink rate.

Results

After processing and analysing the data from the 25 participants to our card guessing game experiment, we did not find any statistical correlation between the average value, standard deviation, or linear interpolation slope of the **eye openness** and the veracity of the participants' answers. We have attempted to analyse the data from various subgroups, depending on experimental conditions such as the nature of the game partner or the participants' predilection to lie, but no correlations were found either in these scenarios.

5.4 Detecting deception using the wireless armband

We also looked at the correlation between the participants' heart rate and the veracity of their answers, as studied in our mock-up crime and interrogation experiment. After studying this physiological manifestation and obtaining valid results using noninvasive measurement devices and methods, but also identifying a series of limitations of the chosen approach, we decided to improve our measurement accuracy by using minimally-invasive measurement techniques.

More specifically, we designed and built a portable, wireless, and ROS-compatible armband capable of measuring the wearer's heart rate and Galvanic Skin Response (GSR) and sending the collected data in real time to a nearby computer. Its technical details are presented at length in section 3.3.2. We asked the participants of the card guessing game, to wear this armband and its sensors on their left hand, wrist, and fingers during the games they played. The armband and its sensors allowed the participants to use their hands freely, being able to choose and pick the cards from the packs during the game.

In this subsection, the analysis procedure used to study the evolution of the participants' heart rate and galvanic skin response during is discussed, as well as the results obtained.

5.4.1 Heart rate

As previously discussed in chapter 2 and also according to our own results from chapter 4, the variations of the human heart rate are one of the main physiological manifestations linked to deception. In our previous research, we have managed to some extent to measure the participants' heart rate by noninvasive means and to find correlations between its variations and the honesty of the participants who underwent our experiments. However, as we wanted to improve the accuracy of our measurement techniques, we decided to monitor it using a minimally invasive device and method: a wireless armband and to analyse its evolution in the context of our new experimental setup.

Measurement technique

The procedure used to measure and analyse the participants' heart rate using the wireless armband was straightforward. The device periodically publishes the last measured heart rate value, refreshed with a frequency of 1 Hz, to a ROS topic to which our main software application subscribes. Every time a new value was received, it was logged automatically to a timestamped CSV file. The data was then analysed post-experimentally in order to search for correlations with the participants' honesty during each round of the game.

Results

We analysed the data obtained from both games played by all the 25 participants to our card guessing game experiment and we obtained two distinct correlations between the **linear interpolation slope** of the participants' heart rate, relative to the reference interval, and the veracity of their answers during each of the game rounds. Each of these correlations was obtained by separating the data according to the game partner experimental condition (human or robot) and analysing each data set separately.

More specifically, the relative **linear interpolation slope** of the participants' **heart rate** had a median value of $\mathbf{-3.39 * 10^{-2}}$ during the rounds when they deceived and $\mathbf{-6.66 * 10^{-2}}$ during those when they were honest with their game partner, with $\boldsymbol{p = 4.31 * 10^{-2}}$, for an analysis done on all the participants, but only on the games played with a **robot** partner. The data was normally distributed and was analysed with the t-test. The participants' heart rate decreased during the card guessing question series and it did so at a slower rate when they deceived their game partner compared to when they were honest to them.

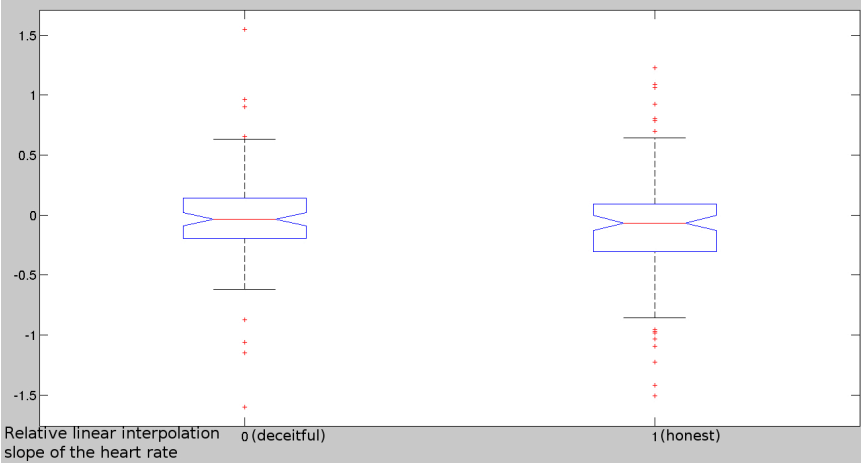

FIGURE 5.5: Relative heart rate linear interpolation slope (robot game partner)

The same parameter (the median value of the relative linear interpolation slope) was also analysed on all the participants, but only on the games played with a **human** partner, obtaining a median value of $\mathbf{-6.65 * 10^{-3}}$ during the rounds when they have deceived and $\mathbf{+2.74 * 10^{-2}}$ during the rounds when they were being honest with their game partner, with $\boldsymbol{F(1, 197) = 4.43}$ and $\boldsymbol{p = 3.66 * 10^{-2}}$. The data was analysed with the ANOVA test. The participants' heart rate decreased during the card guessing question series when they deceived and increased when they were honest with their game partner.

The participants' heart rate evolved differently in the two experimental conditions. In the case of the games played with a human partner their heart rates decreased towards the end of the round and decreased at a higher rate when they were honest to their partner than when they deceived them. However, in the case of the games played with a robot partner the participants' heart rates **increased** during the rounds when they were honest with their game partners, while still decreasing during the rounds where they had a deceptive strategy.

This result, on one hand, seems to somewhat confirm our previous conclusions, presented in chapter 4, where the participants' estimated heart rate were higher when they were honest to their human interrogator than when they deceived them. On the other hand, the contradiction of the results obtained in each of the two experimental conditions implies that there is a significant difference between the participants' physiological manifestations when switching between the two game partners, human or robot. These differences are studied at length in chapter 6.

5.4.2 GSR - Skin conductance

State of the art polygraphs rely heavily on evaluating the galvanic skin response (by measuring the skin conductance or resistance) of the subjects in order to detect whether they are honest or deceptive. Polygraph devices, discussed in chapter 2, which are usually very invasive and use several such skin conductance sensors, placed on various parts of the subjects' bodies. However, simple and portable GSR sensors, that can be placed on the fingers, are also readily available. Moreover, as we have also shown in chapter 2, they have also been successfully used in social robotic experiments.

Measurement technique

We integrated such a sensor, provided by Grove, in our wireless armband. The sensor consisted of a pair of electrodes that were placed on the participants' fingers at the beginning of the experiment, connected by means of a telephone-like coil wire to the main body of the armband. The sensor converted the skin conductance to an electric voltage, whose value, as computed by the analogue-to-digital converter in our armband, was then sent to the computer overseeing the experiment and published into the appropriate ROS topic. This value was then read by our main software application, and then re-converted to obtain the actual value of the skin conductance, measured in microsiemens (μS), using the formula provided by the sensor's manufacturers:

$$G[\mu S] = \frac{512 - V_{digital}}{10 * (1024 + 2 * V_{digital})} \tag{5.1}$$

More specifically, $V_{digital}$ is an integer in the $[0, 1023]$ interval, that corresponds to the ratio between the real value of the output voltage provided by the GSR sensor and the reference voltage of the circuit (typically 5V):

$$V_{digital} = 1023 * \frac{V_{analog}}{V_{ref}}, with V_{ref} = 5V \tag{5.2}$$

Each skin conductance value was logged to a timestamped CSV file which was post-experimentally analysed.

Results

We analysed the data gathered from both games played by all the 25 participants to our card guessing game experiment. No correlation was found between the relative average value, relative standard deviation or relative linear interpolation slope of the participants' skin conductance and the veracity of their answers when analysing the entire data set, containing both games played by each participant. However, we found a correlation between the **relative linear interpolation slope** of the participants' skin conductance with respect to the reference interval in the **robot game partner condition**.

More specifically, the relative **linear interpolation slope** of the participants' **skin conductance** had a median value of $+0.00$ during the rounds when they have deceived and of $+1.93 * 10^{-4}$ during the rounds when they were honest with their game partner, with $p = 2.83 * 10^{-5}$, for an analysis done on all the participants, but only on the games played with a **robot** partner. The data was normally distributed and was analysed with the t-test.

This result means that the participants' skin conductance **increased** (therefore their skin resistance decreased) during the rounds of the card guessing game when they were honest to their robot game partner. An increase in skin conductance (or a decrease in skin resistance) is usually correlated to an increased amount of sweat, which is a significantly better electrical conductor than the skin. However, since the linear interpolation slope defines the variation of this parameter and not its absolute value, the conclusion we can draw is that the amount of peripheral sweating exhibited by the participants was more constant during the rounds when they deceived their robot game partner, while during the rounds when they were honest they tend to sweat more towards the end of the rounds.

5.5 Post-experimental and post-round questionnaires

Some of the main objectives of the new experimental setup were to improve the participants' desire to win the reward, as well as to reduce their perceived stress level. In order to do so, we decided to implement a two-level reward system and we replaced the mock-up crime and interrogation setup in favour of more entertaining experimental format, that was supposed to make the participants feel more at ease even when they were supposed to lie to their interlocutor. These changes also resulted in a series of modifications of the post-experimental questionnaire, that are described in this section.

Moreover, since each game had 8 rounds whose rules and game scenario were identical, and each participant played the game twice, we wanted to understand whether the perception of the participants evolved throughout the game, and also when changing from one game partner to the other. In order to evaluate this evolution, we developed a short post-round questionnaire that the participants had to fill in after each round they play.

5.5.1 Post-experimental questionnaire

Similarly to the previous post-experimental questionnaire, we were interested in evaluating the participants' perspective and the emotions they resented during the experiment. However, we were also interested in evaluating their perception of each of the two different game partners (human and robot), in order to understand whether they perceived them differently or not and if any significant bias may be induced by those differences. We were therefore interested in evaluating the following aspects:

- the participants' perception of their stress level during the game
- their desire to win each of the assigned rewards
- their perception of each game partner
- their performance throughout the game

Taking into account the new objectives, we developed a post-experimental questionnaire containing 12 questions the participants had to answer on a 7-point Likert scale, with 1 corresponding to "Strongly disagree" and 7 to "Strongly agree". The question list is the following:

1. Did you like playing the game?

2. Would you like to play the game again?
3. Did you find it difficult to play the game?
4. Did you want to win the marshmallow or candy?
5. Did you want to get the maximum score, so you can have a chance to win the gift card?
6. Did you find it difficult to answer the questions when you had to tell the truth?
7. Did you find it difficult to answer the questions when you had to lie?
8. Were you frustrated during the rounds when you had to tell the truth?
9. Were you frustrated during the rounds when you had to lie?
10. Were you stressed during the rounds when you had to tell the truth?
11. Were you stressed during the rounds when you had to lie?
12. Do you think your game partner knew when you were deceiving him?

Analysis of results

All the 25 participants also filled in the post-experimental questionnaires we gave them after each of the two games they played. For each of the question, we computed the average value and standard deviation of their answers, with 1 being the minimum value and 7 the maximum value, obtaining the results shown in Table 5.5. The cells coloured in light green indicate which of the two values (for the human or the robot experimental condition) were higher.

TABLE 5.5: The results of the card guessing game post-experimental questionnaire

Question	Human	Robot	Overall
1. Did you like playing the game?	5.72 ±1.21	5.76 ±1.30	5.74 ±1.24
2. Would you like to play the game again?	4.64 ±1.75	4.96 ±1.84	4.80 ±1.78
3. Did you find it difficult to play the game?	2.08 ±1.26	2.76 ±1.48	2.42 ±1.40
4. Did you want to win the marshmallow or candy?	5.20 ±1.91	5.08 ±2.06	5.14 ±1.97
5. Did you want to get the maximum score, so you can have a chance to win the gift card?	5.72 ±1.67	5.88 ±1.67	5.80 ±1.65
6. Did you find it difficult to answer the questions when you had to tell the truth?	2.28 ±1.40	2.48 ±1.53	2.38 ±1.46
7. Did you find it difficult to answer the questions when you had to lie?	3.20 ±2.02	2.80 ±1.61	3.00 ±1.82
8. Were you frustrated during the rounds when you had to tell the truth?	1.96 ±1.54	2.00 ±1.38	1.98 ±1.45
9. Were you frustrated during the rounds when you had to lie?	2.68 ±2.04	2.44 ±1.69	2.56 ±1.85
10. Were you stressed during the rounds when you had to tell the truth?	2.24 ±1.54	2.88 ±1.94	2.56 ±1.76
11. Were you stressed during the rounds when you had to lie?	3.72 ±2.11	3.64 ±1.87	3.68 ±1.97
12. Do you think your game partner knew when you were deceiving him?	3.20 ±1.94	2.80 ±1.66	3.00 ±1.80

As each participant played the game twice, once with a robot partner and once with a human partner, they answered the questionnaire twice, once after each game. We can observe that they liked playing the game regardless of the game partner they had (H = 5.72, R = 5.76) and that they wanted to play the game again, albeit slightly more after playing the game with a robot for the first time (H = 4.64, R = 4.96). Their overall desire to win the first reward (the marshmallow or candy) was slightly lower than in the case of the first experiment (5.14 compared to 5.37 previously), but their desire to achieve the maximum score and therefore to have a chance to win the gift card was significantly higher (5.80 ±1.65).

In this post-experimental questionnaire, we were also interested in comparing the participants' perception during the rounds when they had to tell the truth and the rounds when they were supposed to deceive their game partner. We observed that, as expected, during the rounds when they were expected to be deceitful they

experienced higher level of frustration (2.56 vs. 1.98) and stress (3.68 vs. 2.56), and also they considered answering the questions more difficult (3.00 vs 2.38). Nonetheless, the overall level of perceived frustration, stress and difficulty was under average (less than 3.5), therefore being an improvement with respect to the first experiment.

When comparing their perception of the games played with a robot partner to those played with a human partner, we can also observe that they felt slightly less stressed when they had to lie to a robot (H = 3.72, R = 3.64), yet more stressed when they had to tell the truth to a robot (H = 2.24, R = 2.88). A similar trend was observed when comparing the perceived difficulty, lower when having to deceive a robot partner (H = 3.20, R = 2.80) and higher when having to tell the truth to a robot partner (H = 2.28, R = 2.48). This tendency can be easily explained by the fact that, rather consistently with their perception of the last experiment, they considered the human game partner more capable of knowing when they were deceiving them (H = 3.20, R = 2.80).

5.5.2 Post-round questionnaire

Last, but not least, we analysed the evolution of the participants' perceived stress, frustration and difficulty levels from one round to another, as they get used to playing the game. In order to establish that, we asked them to fill a short, 3-question, post-round questionnaire at the end of each game round:

1. Did you find it difficult to answer the questions of this round?
2. Did you feel frustrated during this round?
3. Did you feel stressed during this round?

Analysis of results

If the frustration level was almost constant from one round to the other, with no relevant variations, both the perceived difficulty and stress level showed a significant decrease towards the end of the game when playing the game with a human partner, with the exception of the last rounds when they had to be honest to a robot partner, when the participants reported an increase of their perceived stress and difficulty level with respect to the previous round.

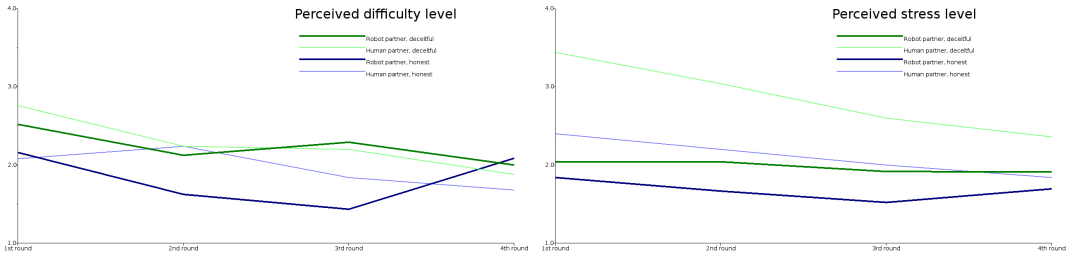

FIGURE 5.6: Perceived difficulty and stress level evolution

The thin lines show the evolution of the perceived level when playing the game with a human partner and the thick ones correspond to the games played with a robot partner. The green lines are associated to the rounds when the player had to be deceitful, while the blue ones are associated to the rounds when the player had to be honest.

5.6 Summary

In this chapter, we discussed our attempts to detect deception using non-invasive sensors, such as an RGB camera, in conjunction with minimally-invasive sensors, such as the wireless armband we have designed. We have detailed the experimental methodology, the data measurement and analysis procedures used to collect the raw physiological measurements, as well as the results of our post-experimental analysis. In particular, we analysed the participants' head position and orientation, their eye openness, their heart rate, as well as their skin conductance. We attempted to find correlations between the evolution of these parameters during each of the 8 rounds of our card guessing game and the veracity of their answers during each of these rounds, depending on the colour of the card they picked.

Moreover, this exact experimental setup and collected data also allowed us to study the differences between the participants' behaviour when playing the game with a robot partner compared to playing the same game with a human partner. These differences are discussed at length in the next chapter.

5.6.1 Results

We found a series correlations between the measured head height (which is inversely proportional to the interaction distance), head vertical position, the head roll, the heart rate and the skin conductance of the participants and the veracity of their answers during each round of the game. The results can be summarised in Table 5.6, with the light green cells indicating which of the two median values (the one corresponding to the rounds when participants were expected to tell the truth and the other one corresponding to the rounds when participants were expected to lie) was higher, for each of the monitored parameters and measures.

TABLE 5.6: Overall results

Parameter	Measure	$\tilde{x}$: lie	$\tilde{x}$: truth	EPQ-L	Partner	Conclusion
head height	slope	$5.12 * 10^{-2}$	$2.67 * 10^{-2}$	all	both	increasing at a higher rate when lying
head height	slope	$7.26 * 10^{-2}$	$1.26 * 10^{-2}$	all	human	increasing at a higher rate when lying
head height	slope	$6.86 * 10^{-2}$	$1.81 * 10^{-2}$	high	both	increasing at a higher rate when lying
head **y** *(inv. axis)*	average	$-7.75 * 10^{-1}$	$3.78 * 10^{-1}$	all	robot	higher position when lying *(inv. y axis)*
head roll	slope	$-1.61 * 10^{-3}$	$-2.38 * 10^{-3}$	low	both	clockwise increase, at a lower rate when lying
heart rate	slope	$-3.39 * 10^{-2}$	$-6.66 * 10^{-2}$	all	robot	decreasing at a lower rate when lying
heart rate	slope	$-6.65 * 10^{-3}$	$2.74 * 10^{-2}$	all	human	decreasing when lying, increasing when honest
skin conductance	slope	0.00	$1.93 * 10^{-4}$	all	robot	increasing when honest

These results have also been published in a conference article entitled *Detecting deception in HRI using minimally-invasive and noninvasive techniques*, to the the RO-MAN 2019 conference (14-18 October 2019, New Delhi, India).

Moreover, the analysis of the feedback received from our participants showed that our experimental design was an improvement over the previous one, discussed in chapter 4 as the participants found they were even more enticed to obtain their rewards than in the case of the previous scenario, and as their perceived stress level was lower than in the case of the interrogation part of the previous experiment.

5.6.2 Discussion

Compared to the first part of our work, presented in the previous chapter, the results of this experiment represented a significant improvement. Part of the positive evolution of this research was due to the improved noninvasive measurement devices

and techniques used in this setup, while another significant contribution came from the use of the minimally-invasive armband. These results have been obtained in an in-the-wild experimental setup, where participants played a card game and had no behavioural or mobility constraints imposed by the experimenter. Therefore, we can safely assume that these results can be applied to a larger variety of interaction scenarios.

However, a practical point should be made concerning the use of the minimally-invasive measurement devices. Even if these devices do not significantly interfere with the user's ability to perform everyday tasks and can be worn or carried around without being a burden, they can still be removed by the user if they desire so. As situations where the user wants to interfere with the assistive robot's ability to detect deception may arise, one of the options that the human user can always have is to remove any wearable or portable measurement devices they are wearing. This means that even though we have proved that the use of minimally-invasive measurement devices can improve a robot's ability to detect deception, its human user has the ability to voluntarily limit the robot's lie detection abilities. Therefore, if such devices are used in real life HRI scenarios, measures must be taken in order to at least detect when the human users are interfering with the measurement devices and react accordingly. This is significantly less of an issue in the case of noninvasive sensors, such as cameras embedded in the robot's body.

Last, but not least, the correlations we found and especially the large variety of physiological manifestations that we proved to be correlated in various ways to the participants' honesty could definitely be used to train a predicting system that would attempt to estimate to what extent the participants are honest in real time. Moreover, once the predictor is trained, it could function in real time, as all the data is collected in real time as well and no frequency-domain analysis is performed. This would be the most obvious future development of our current work, but it would require a significantly higher number of samples for training and validating the system. If we consider that all the data collected for one round played by one player represents one sample, we have collected roughly 400 valid samples (approximately 200 for each type of game rounds, honest and deceptive).

5.7 My contribution

The main contributions to this chapter were in the design and implementation of the experimental setup (including the card game rules, card layout and question trees), the software used to collect and analyse the physiological parameters studied throughout the experiment, as well as the post-experimental questionnaire used to evaluate the participants' experience during the experiment.

Chapter 6

Comparing human-robot and human-human interactions

6.1 Introduction

The main part of our research work focuses on the study of the physiological manifestations that can be correlated to deception in HRI. This study is based on the previous findings found by psychologists, psychoanalysts, and psychiatrists when studying the same issue in inter-human interactions, which we have been discussed at length in chapter 2. These studies and findings are indeed a major source of inspiration for our work, but, as we have concluded in section 2.5, there is no universal proof that their results are also implicitly valid for HRI scenarios, nor that the behaviours associated to deception in HRI are similar to those exhibited in inter-human interactions. Since there is no such proof, our hypothesis is that we have to take into account the nature of the game partner, human or robot, when attempting to detect deception and also that we cannot automatically apply the previous findings concerning deception detection in inter-human interactions to the domain of HRI.

We also decided to investigate the differences between inter-human interactions and HRI. More specifically, as we designed an experimental setup that helped us study the manifestations associated to deception, we made sure that each participant undergoes the same experiment twice, interacting once with a human partner and once with a robot partner. This would therefore allow us to contrast and compare the exact same physiological manifestations monitored in order to detect deception while taking into account the new experimental condition: the nature of the game partner.

This chapter starts with a quick review of the experimental setup used and the physiological parameters monitored for this analysis, as detailed in Chapter chapter 5. We continue by presenting the results of the analysis we have done of the impact of the nature of the interlocutor on the physiological manifestations exhibited by the participants, including the evaluation of the post-experimental and post-round questionnaires. Then, we discuss the meaning and importance of these results, as well as whether they confirm or not if we can automatically use in HRI the conclusions already drawn when studying deception in inter-human interactions. This chapter ends with a sum up of my contribution to the field.

6.2 Experimental setup and evaluated parameters

The experimental setup we used and the physiological parameters monitored were the same as for the experiment detailed in the previous chapter, in section 5.2. The

experiment was designed from the very beginning so as it would allow us to study on one hand the physiological mechanisms of deception in HRI, as well as the differences between the interactions undergone with a robot and those undergone with a human in this exact scenario. In this section, we simply review the main elements of the experimental setup we used, as well as the list of physiological parameters, devices, and methods used for this study.

6.2.1 Experimental setup review: the card guessing game

The experiment took the form of a card guessing game played between the participants and a game partner, who was either human or robot. At each of the 8 game rounds, the participant chose an image card from a deck of 32 cards and drew a small card containing the main round objective: either mislead or help their game partner guess the image card they had drawn. Then, their game partner asked them 5 questions, trying to guess the chosen card. At the end of each round, if the participants achieved their goal of helping or misleading their game partner, they won a point. If they score 4 points or more, they were awarded some candy, while if they achieved the maximum score of 8 points, they got a chance to win a 50€ gift card.

Each of the 25 participants played the game twice, once with a human partner and the other time with a robot partner. In order to reduce to a minimum the bias induced by the order in which they played the two games, we divided the participants into two groups, one consisting of people who play their first game with a human partner and the other consisting of people who play their first game with a robot partner. They were evenly split depending on their gender (with 2 women in each of the two groups), as well as their predilection to lie, measured using the short, revised version of the EPQ-L questionnaire. The average value of the EPQ-L was of 6.42 in the case of the group who played their first game with a robot partner and of 6.15 in the case of the other group.

The data annotation was done in a similar fashion as previously described in chapter 5 , identifying one baseline interval and 8 round intervals, with the truth value of the participants' statements being established based on the colour of their chosen card and whether they have achieved their objective during that round. However, there was one major difference with respect to the timestamps of each measurement interval. If in the previous analysis, we focused on analysing the data acquired between the moment when the game partner started asking the questions necessary to guess the card and the moment the player showed their game partner the two cards, in this case we extended the length of these intervals, as we were interested in monitoring the participants' behaviour over a longer time frame. Particularly, in the case of this analysis, the beginning of each interval corresponded to the moment when the participant was asked to choose a card from the deck of image cards and ended when they were asked to show the two cards.

6.2.2 Evaluated parameters and analysis methods

The parameters monitored in order to detect deception in chapter 5 were the following:

- head position and orientation
  - x,y position
  - height and width
  - roll, pitch and yaw angles

- heart rate
- skin conductance

The head position and orientation were evaluated non-invasively, using a Logitech C920 high resolution camera, while the heart rate and skin conductance were measured using the wireless armband that we developed. These parameters were tracked in real time by our main software application, logged to individual timestamped CSV files, and then analysed post-experimentally.

However, the post-experimental analysis was made using different experimental conditions. As previously, we computed, for each game round interval, the relative average value, standard deviation, and slope of linear interpolation of the data relative to those of the reference interval. Nonetheless, when statistically analysing the entire data sets, we attempted to correlate these measures to the **nature of the game partner** instead of doing so to the veracity of the participants answers.

6.3 Results

We analysed the data gathered from both games played by all the 25 participants to our card guessing game experiment. We computed the relative average values, standard deviations, and slopes of the linear interpolation for each game round relative to the reference interval and then compared the values from the two sets of games: the ones undergone with a human game partner and the ones undergone with a robot game partner. We found correlations between the evolution of several measures of the **head position and orientation**, **skin conductance**, and the **eye openness** of the participants and the nature of their game partner, that are resumed in Table Table 6.1.

TABLE 6.1: Analysed measures vs. nature of the game partner

Parameter	Measure	Test	p-value	$\tilde{H}$	$\tilde{R}$
Head x	std. dev.	t-test	$\mathbf{1.02 * 10^{-12}}$	$\mathbf{2.71 * 10^{-1}}$	$\mathbf{4.90 * 10^{-1}}$
Head y	std. dev.	t-test	$\mathbf{3.61 * 10^{-16}}$	$\mathbf{1.32 * 10^{-1}}$	$\mathbf{2.27 * 10^{-1}}$
Head width	std. dev.	Krusk.-Wall.	$\mathbf{4.01 * 10^{-9}}$	$\mathbf{5.11 * 10^{-1}}$	$\mathbf{4.02 * 10^{-2}}$
Head height	std. dev.	ANOVA	$\mathbf{7.58 * 10^{-6}}$	$\mathbf{4.97 * 10^{-1}}$	$\mathbf{5.06 * 10^{-2}}$
Head roll	average	ANOVA	$\mathbf{1.31 * 10^{-2}}$	$\mathbf{3.05 * 10^{-2}}$	$\mathbf{1.12 * 10^{-2}}$
Head pan	std. dev.	ANOVA	$\mathbf{3.33 * 10^{-6}}$	$\mathbf{-2.48 * 10^{-2}}$	$\mathbf{3.34 * 10^{-2}}$
Head tilt	std. dev.	ANOVA	$\mathbf{1.31 * 10^{-4}}$	$\mathbf{-8.01 * 10^{-3}}$	$\mathbf{4.17 * 10^{-2}}$
Skin conductance	average	t-test	$\mathbf{2.81 * 10^{-22}}$	$\mathbf{2.16 * 10^{-1}}$	$\mathbf{2.04 * 10^{-1}}$
	slope	t-test	$\mathbf{3.48 * 10^{-6}}$	$\mathbf{7.51 * 10^{-4}}$	$\mathbf{2.20 * 10^{-5}}$
Eye openness	average	t-test	$\mathbf{1.63 * 10^{-3}}$	$\mathbf{-1.15 * 10^{-2}}$	$\mathbf{-1.34 * 10^{-2}}$
	std. dev.	ANOVA	$\mathbf{1.42 * 10^{-2}}$	$\mathbf{7.80 * 10^{-3}}$	$\mathbf{2.25 * 10^{-3}}$

The data sets tested with the ANOVA test and the t-test were normally distributed, while the data set tested with the Kruskal-Wallis test was not normally distributed. The $\tilde{H}$ and $\tilde{R}$ columns display the median value of the given measure in the human and robot experimental conditions respectively, while the cells coloured in light green indicate which of the two median values (from the robot or human condition) was higher. No statistically significant correlations were found between the participants' heart rate and the nature of the game partner.

Head position and orientation

Firstly, the results we obtained indicate that the participants **moved their heads more** in the camera plane, both on the horizontal and on the vertical axis, when they played the game **with a robot game partner** than with a human game partner. This behaviour is also confirmed by the analysis of the head pan and tilt movements. The participants also varied their head pan and tilt **more in the robot game partner condition** than in the games played with a human partner. If the first pair of results (x and y head position) define the position in the camera plane, the second set of results describe the left-to-right and top-to-bottom rotational movements of the head.

However, their face width and height, which is inversely proportional to the interaction distance, **varied less** when they played the game **with a robot game partner** than with a human game partner, which means that their interaction distance also varied less in the human game partner condition. A more constant interaction distance is a usually associated with a more controlled upper body posture when sitting down at the table.

Another observed behaviour was the fact that the average head roll angle was **higher when playing the game with a human partner** than it was when playing the game with a robot partner, relative to the reference interval, which means that the participants' head was rotated more in a clockwise direction when the participants played the game with a human partner.

Skin conductance

Skin conductance, which is inversely proportional to the skin resistance, is mainly influenced by the amount of sweating the participants exhibit. A higher skin conductance usually implies a higher level of sweating, due to the good electrical conductance properties of sweat. As we have observed in our research, our participants exhibited a **higher skin conductance** when played the game against a human game partner than when played the game against a robot game partner, which implies that they sweated more in the **human experimental condition**.

On top of that, the slope of the linear interpolation of the skin conductance was also **higher in the human game partner condition**, which means that the skin conductance also **increased more** during each round played with a human partner.

Eye openness

Last, but not least, we have also obtained a series of correlations between the participants' eye openness and the nature of the game partner. On one hand, the participants **blinked more** when playing the game with a **human game partner** than with a robot game partner, as the value of the standard deviation of the eye openness was higher in the **human** game partner condition. Moreover, the average value of the eye openness was also **higher in the human game partner condition**, meaning that the participants eyelids were more open in average when playing with a human than when playing with a robot.

6.3.1 Discussion

In the previous chapter, we have studied the differences between the physiological and behavioural manifestations exhibited by the participants during the rounds when the participants deceived their game partner in a card guessing game and

those when the participants were honest to their game partner. We have proved the existence of several such differences, concerning the **head position and orientation**, the **heart rate** as well as the **skin conductance** of the participants. The participants were studied during a card guessing game experimental setup that involved playing the same card guessing game twice, once against a robot and once against a human partner.

In this chapter, we have shown the existence of a similar series of differences regarding the physiological and behavioural manifestations exhibited by the participants, this time between the card guessing games played against a human partner and those played against a robot partner. These differences concerned the **head position and orientation**, the **skin conductance** and the **eye openness** of the participants. The data used for the analysis in this chapter is the exact same data collected in the experiment defined in chapter 5, the only difference being the perspective of the analysis.

First of all, the fact that we found that both the **game partner condition** and the **answer honesty condition** have an impact on the studied parameters shows already that the nature of the interlocutor must be taken into account when attempting to detect deception using a given set of monitored parameters and measurement techniques. Moreover, even if we can use the findings from the previous studies of deception in inter-human interactions as starting points for deception detection in HRI, we cannot assume that a system trained to detect deception in inter-human interactions will automatically detect deception in HRI and vice-versa.

Indeed, some of the parameters whose evolution was correlated to one of the experimental conditions did not yield a similar conclusion with respect to the other experimental condition, but Table 6.2 showed the list of parameters whose evolution was correlated to both experimental conditions:

TABLE 6.2: Analysed measures: comparison between the honesty condition and the game partner condition

Parameter	Measure	Game partner		Round honesty		Experimental conditions
		$\tilde{x}_{human}$	$\tilde{x}_{robot}$	$\tilde{x}_{truth}$	$\tilde{x}_{lie}$	
Head **y** coordinate	std. dev.	$1.32 * 10^{-1}$	$2.27 * 10^{-1}$			all participants
	average			$-7.75 * 10^{-1}$	$3.78 * 10^{-1}$	robot partner only
Head height	std. dev.	$4.97 * 10^{-1}$	$5.06 * 10^{-2}$			all participants
	slope			$5.12 * 10^{-2}$	$2.67 * 10^{-2}$	
Head roll	average	$3.05 * 10^{-2}$	$1.12 * 10^{-2}$			all participants
	slope			$-1.61 * 10^{-3}$	$-2.38 * 10^{-3}$	low EPQ-L
Skin conductance	average	$2.16 * 10^{-1}$	$2.04 * 10^{-1}$			all participants
	slope	$7.51 * 10^{-4}$	$2.20 * 10^{-5}$			
	slope			0.00	$1.93 * 10^{-4}$	robot partner only

As this table shows, using the results previously detailed in chapter 5 and chapter 6, the participants' **head vertical position**, **head height** (dependent on the interaction distance), **head roll angle** and **skin conductance** were both correlated to the nature of the game partner and the honesty of the participants during each of the game rounds. The light green cells indicate in which of the game partner condition cases the analysed measure was higher (human or robot partner), while the light blue cells indicate in which of the round honesty condition cases (honest or deceitful rounds).

However, this analyse does not allow us to quantify the exact impact of each of these two experimental conditions to the ability to detect deception, especially since the **measures** that yielded correlations in each of the two experimental conditions

were different for each monitored parameter. Even in the case of the **skin conductance linear interpolation slope**, whose evolution was correlated both to the nature of the game partner and the honesty of the participants, the latter was observed only when studying the players who played the game against a robot game partner. Future work on this topic could potentially allow us to understand exactly to what extent each of the parameters has an impact on the ability to detect deception.

6.4 Post-experimental questionnaires

The structure, the content, and the results of the post-experimental questionnaires are the same as those presented in section 5.5. They were design in such a way that would also allow us to compare the participants' perspective of the experiment so as to analyse the differences between their perception when playing the game with a robot and when playing the game with a human.

In particular, we are interested in understanding if there were any differences between their perceived difficulty, frustration, and stress levels, as well as their perception with respect to the ability of their game partner to know when they were deceived. We already presented these results in the previous chapter, but here they analyse in more details from the perspective of the game partner condition. The **GP** column identifies the game partner, which is either human (**H**) or robot (**R**), while the other columns indicate the average value of the answer (**Avg**) and the number of each answers (on the columns marked **1** to **7**) of each question.

TABLE 6.3: Post-experimental questionnaire answers: average and distribution

Question	GP	Avg	1	2	3	4	5	6	7
Did you find it difficult to answer the questions when you had to **tell the truth**?	**H**	**2.28**	**8**	**11**	1	1	4	0	0
	R	**2.48**	**7**	**11**	1	0	6	0	0
Did you find it difficult to answer the questions when you had to **lie**?	**H**	**3.20**	5	**7**	**6**	0	1	4	2
	R	**2.80**	**6**	**7**	5	2	3	2	0
overall perception of the game difficulty	**H**	**2.74**	-						
	R	**2.64**	-						
Were you frustrated during the rounds when you had to **tell the truth**?	**H**	**1.96**	**14**	**6**	1	2	1	0	1
	R	**2.00**	**13**	**6**	2	1	3	0	0
Were you frustrated during the rounds when you had to **lie**?	**H**	**2.68**	**10**	**7**	0	3	0	4	1
	R	**2.44**	**10**	**7**	1	3	2	2	0
overall perception of frustration	**H**	**2.20**	-						
	R	**2.22**	-						
Were you stressed during the rounds when you had to **tell the truth**?	**H**	**2.24**	**11**	**7**	2	0	5	0	0
	R	**2.88**	**9**	**4**	3	3	**4**	0	2
Were you stressed during the rounds when you had to **lie**?	**H**	**3.72**	**5**	**5**	2	2	4	**5**	2
	R	**3.64**	4	4	**5**	2	**5**	4	1
overall perception of stress	**H**	**2.98**	-						
	R	**3.26**	-						
Do you think your game partner knew when you were deceiving him?	**H**	**3.20**	**5**	**7**	4	2	3	2	2
	R	**2.80**	**7**	**6**	3	5	3	0	1

Analysis of results

One of the first results that can be observed is that the participants believed the **human** game partner was more able to detect when they were deceiving him/her than the robot was, even though overall both the human partner and the robot partner's ability to detect deception was considered under average, with average answer values of 3.20 and 2.80, respectively. Moreover, almost half of the participants (12 and 13, respectively) considered strongly disagreed or disagreed that their game partner was able to know when they were deceiving him.

However, the overall perception of the **stress level** was higher when playing the game against the robot (3.26 vs. 2.98), especially in the case of the rounds when they were expected to tell the truth (2.88 vs. 2.24). We find it interesting and somewhat contradictory that participants were more stressed when playing the game with the game partner who was considered less able to detect deception.

Regardless, the perceived difficulty level seems to partly justify both these results. The perceived difficulty level was higher in the case of the **human partner** when participants were expected to lie, with values of 3.20 for the human partner and 2.80 for the robot partner, which are in line with the previously presented values concerning the expected ability to detect deception. This shows a correlation between the overall perceived difficulty during the rounds **when participants were expected to lie** and their overall perception with respect to each partner's ability to detect deception.

On the other hand, the participants felt that it was more difficult to play **with a robot game partner** during the rounds when they were expected to tell the truth. This conclusion is in line with their perceived stress level in the same conditions, which was also higher with a robot game partner. This shows a correlation between the overall perceived stress level during the rounds **when participants were expected to tell the truth** and their overall perception of the difficulty level of those rounds.

These results have also been published in a conference article entitled *The impact of a robot game partner when studying deception during a card game*, to the International Conference on Social Robotics (ICSR 2019), (26-29 November 2019, Madrid, Spain).

6.5 Summary

In this chapter, we analysed the impact of the nature of the interlocutor on the physiological and behavioural manifestations of the participants to our experiment. The experimental setup used was the exact same one detailed in chapter 5: a card guessing game scenario where participants are enticed to lie during half of the game rounds and played twice by each participant, once with a human partner and once with a robot partner. The analysis was done on the exact same data sets collected previously.

We studied several parameters: the **head position and orientation**, the **eye openness**, the **skin conductance**, and the **heart rate** of the participants. We showed the existence of a series of correlations between the nature of the game partner (human or robot) and the participants' **head position and orientation**, **skin conductance**, and **eye openness**. Moreover, after analysing the post-experimental questionnaires given to the participants, we have found some differences between their perception of the games played with each of the two game partners.

If we sum up the results of the previous and the current analysis, we have found that the participants **head vertical position**, **head height** (dependent on the interaction distance), **head roll**, and **skin conductance** are correlated both to the nature of the interlocutor and the veracity of the participants' answers.

These findings show that our hypothesis for this part of our work, which was that the nature of the interlocutor (human or robot) must be taken into account when attempting to deception **was validated**. Implicitly, we also showed that we cannot automatically use the previous findings of the research done in deception detection in inter-human interactions to deception detection in HRI, even though these findings still represent a cornerstone to our research. However, we were not able to quantify

exactly in what ways we should take into account the nature of the interlocutor to improve the ability of a robot to detect deception.

6.6 My contribution

The main contributions to this chapter were mainly the design and implementation of the experimental setup (including the card game rules, card layout and question trees), the software used to collect and analyse the physiological parameters studied throughout the experiment, as well as the post-experimental questionnaire used to evaluate the participants' experience during the experiment.

Chapter 7

Detecting deception using audio analysis

7.1 Introduction

In the previous chapters, we focused on analysing aspects of the behaviour, as well as the physiological manifestations that can be associated to deception, using non-invasive and minimally-invasive devices and analysis techniques. Therefore, in these previous chapters, we focused on techniques that follow the guidelines of the **polygraph approach**.

However, we have not yet studied nor implemented any deception detection techniques that follow the guidelines of **cognitive load** based approaches. The main premise of these techniques, as discussed in chapter 2, is that making false statements or giving deceitful answers to comments **increases the cognitive load** of the deceptive person compared to a scenario where that person is honest in their answers or statements. This increase of the cognitive load can be evaluated in several ways, one of them being the analysis of the participants' **response times** to questions.

The analysis of response times has already been used as a deception detection technique in both inter-human interactions [20] [21] [53], and HRI [41] [42]. The results were similar in both situations, indicating that the response times were **higher** when the subjects were being deceitful than when they were being honest. However, in most of the previous studies on this topic, the response times were manually measured and extracted using audio recording and replaying software. Since we want to develop detection techniques that can be easily integrated into automated HRI scenarios, one of our objectives was to develop a technique that allowed us to measure the response times automatically.

Using the exact same card guessing game experimental setup as the one described at length in chapter 5, we decided to analyse the **response times** of the participants when answering the questions they were asked during the guessing phase of the game. We used the audio recordings collected throughout the experiments and developed an automated response time extraction algorithm, using the previously-made manual annotations of the game intervals as a guide and we used a similar analysis methodology as that used in chapter 5. Our current response time extraction algorithm was run post-experimentally, but it can be easily adapted to work in real time.

Even if this part of the thesis can be easily included in chapter 5, we decided to dedicate a separate chapter for it as it analyses a completely different component of the interaction than what we discussed in the previous chapters, also using premises of a completely different category of deception detection techniques. In this chapter, we present the analysis methodology and algorithm used to automatically extract

the response times for each participant and continue by showing and discussing the results of our analysis.

7.2 Experimental setup and data analysis

A major part of the cognitive load based deception detection focuses on the development of specific interaction and interrogation scenarios that would emphasise the increase of the subjects' cognitive load when they are deceitful and that minimise the possibility of them having predefined deceitful answers to their questions. However, as mentioned previously, the experimental setup we have used for this part of our research is the exact same one used for the experiment we have detailed in the previous chapters. Therefore, we did not implement a specific scenario that would maximise our chances to detect deception using this particular detection technique.

In this chapter, we focused only on the analysis of the audio samples recorded using the Logitech C920 RGB camera. Therefore, in this section, we start by reviewing the main elements of the experimental setup we used, and then continue by detailing the analysis procedure employed to extract the response times from the recorded audio samples.

7.2.1 Audio analysis procedure: extracting the response times

For each game played by the participant, we recorded an audio mp3 file using the Audacity open-source software and the microphone input provided by the Logitech C920 RGB camera. Each file was then post-experimentally analysed, using a Scilab script that relied on the manual annotations done by the experimenter.

Manual annotations register the key moments of the game (beginning and end of each game round), as well as whether the participant was deceitful or honest during each round. These annotations were used, on one hand, for data segmentation, as the manually annotated timestamps were necessary to identify each question interval. On the other hand, the measured response times were correlated to the objective of the participant during each round, established using the manual annotations of the veracity of their answers during that round, based on the chosen cue card (red or green). All the data was annotated by one of the authors, while the recordings of 4 random participants have been chosen for annotation by a second rater. The Cohen's kappa coefficient for inter-rater agreement had a value of $\boldsymbol{\kappa = 0.938}$, which corresponds to almost perfect agreement.

In this case, for a given participant, we only defined 8 analysis intervals, one for each round of the game. The reference interval was not relevant for this analysis, since there were no questions asked to the participant during that time frame, the game partner being the only one talking. The round intervals were measured, as it was the case in Chapter chapter 5, between the moment when the participant was asked the first card guessing question and ended when they were asked to show the two cards. Each of these intervals begun with the 5 question and answer pairs regarding the chosen card, that the participant was expected to answer honestly or dishonestly depending on the chosen cue card (green or red). The response times we extracted, were the intervals passed between the end of each question and the beginning of each answer.

Using the manually annotated timestamps to roughly identify and cut the questioning interval for each game round, the Scilab analysis script we developed, processed the audio samples in 4 stages:

1. Thresholding
2. Continuous pause extraction
3. Valid pause selection
4. Response time extraction

In the first stage, the waveform of the audio sample was thresholded as follows. First, the mean absolute value of the audio level of the entire interval was computed. Then, the audio samples where the absolute value of the recorded audio level was **lower than 25%** of the previously computed mean absolute level were **set to 0**, using a fixed size moving window. The moving window had a length of **250 ms** and its measurement step was of **50 ms**. Therefore, for a "silent sample" to be detected by the algorithm, it had to be longer than 250 ms, while the pause extraction algorithm resolution was equal to the step of the moving window (50 ms). This was done so as to ignore the small pauses that may occur in the discourse of each participant and to detect only the moments of silence that separate the dialogue of the two interlocutors, while still ensuring a good detection resolution.

Then, using the thresholded "silent" samples, we proceeded to identifying continuous pauses, also computing the start and end time for each of these continuous pauses. In particular, we searched for continuous windows of samples whose values had been set to 0 in the previous stage of the processing. If the audio sample that was being analysed, corresponding to one game round, ended with a silent interval, the last such silent interval was discarded, as we cannot compute its exact end time. Afterwards, the **first**, **third**, **fifth**, **seventh**, and **ninth** such pauses of each game interval were selected, as these pauses corresponded to the pauses between each question asked by the game partner and each answer given by the participant.

Lastly, each of the 5 response times was computed based on the start timestamps and end timestamps of each valid silent interval, that had been previously extracted.

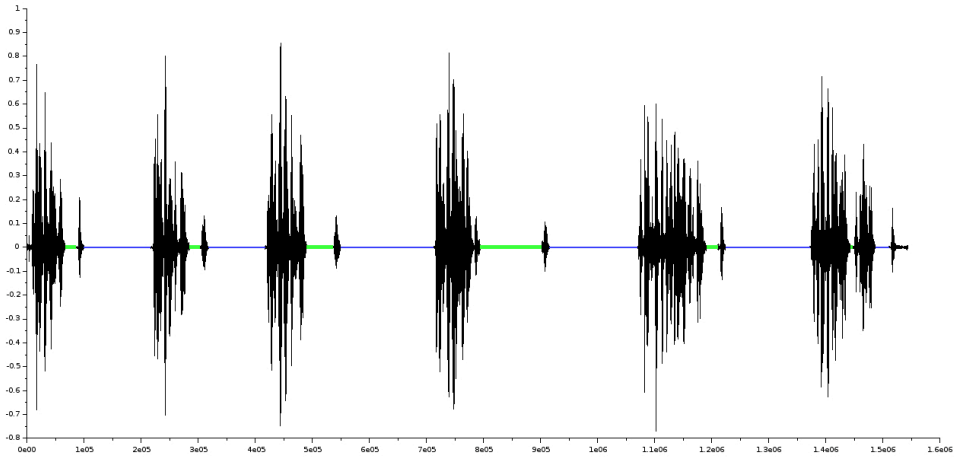

FIGURE 7.1: Display of the response time extraction algorithm

In Figure 7.1, we displayed a graphical example of the processing done by our response time extraction algorithm. The graph represents the audio sample recorded from one participant during one game round, that had been pre-cut using the manual annotations of the start and the end of the round. This audio sample was then analysed by our algorithm, and the pauses and the response times we are interested in had been identified automatically. The **green** horizontal lines show the relevant

pauses that have been extracted (those that correspond to the response times we are monitoring), while the **blue** horizontal lines show other pauses that are detected, but are not taken into account in the analysis, as they do not correspond to the response time of participants to their interlocutor's questions.

Lastly, we merged the data collected from all participants, as we have done previously, with the difference that in this case the **absolute values** of the response times are being logged instead of the ones **relative to the reference interval**. We therefore created a CSV file containing on each line pairs of values, the first one corresponding to one extracted response time and the second corresponding to the assigned truth value of the participants' answers during that interval. If a participant had drawn a green card and won a point, the participant was considered honest during the entire interval. Otherwise, if a participant had drawn a red card and won a point, they were considered deceitful during that interval, regardless of how many of their answers were deceitful. This approach, similar to the one used for the analysis we have done in chapter 5, was chosen as we are interested in the participants' overall objective during a given round and not the veracity of each of their individual answers. In the rare occurrence of a participant not winning a point during that round, the data was discarded.

7.3 Results

We analysed the data collected from all the 50 games, attempting to search for correlations between the measured response time and the veracity of the participants' answers during a given round. We observed the following correlation: the participants **took more time to answer** during the rounds when they were expected to be **deceptive** (the rounds when a red card was drawn) than during the rounds when they were expected to be honest.

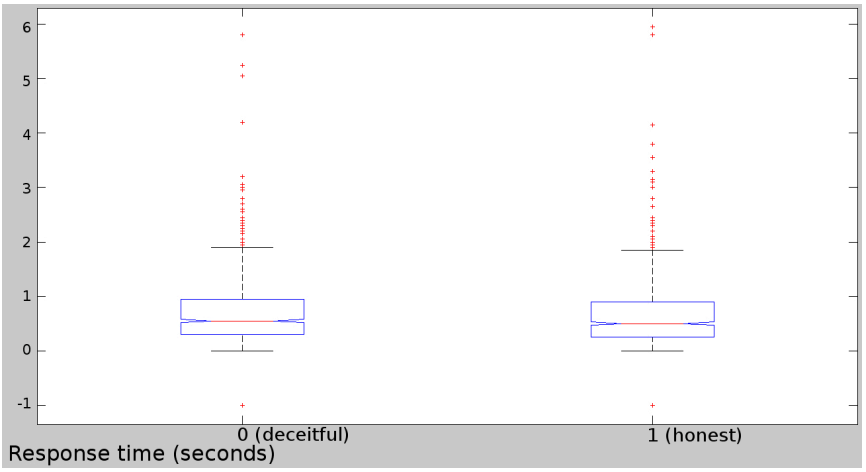

FIGURE 7.2: The response time to questions (seconds)

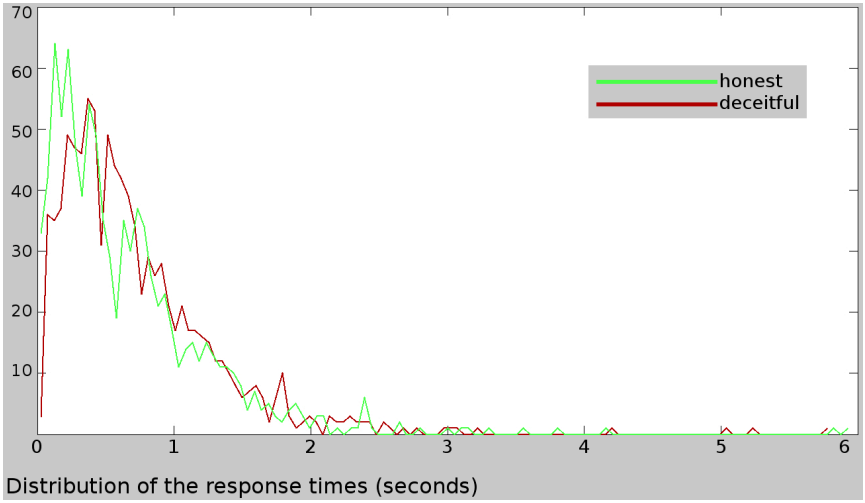

FIGURE 7.3: The distribution of the response time to questions (seconds)

In particular, the median value of the participants' response time was of **550 ms** during the rounds when they **deceived** their game partner and of **500 ms** during the rounds when they were **honest** to their game partner, with **p = 0.041**. The data was **normally distributed** and was analysed with the **Wilcoxon rank sum test**. The results are graphically detailed in Figure 7.2 and Figure 7.3. This result confirms previous results from both inter-human and human-robot interaction scenarios, where the participants' response times were also higher when being deceptive than when telling the truth.

7.4 Summary

In this chapter, we studied the evolution of the response time of the participants to the questions asked during a card guessing game scenario that involved deception, the same scenario used for the research detailed in the two previous chapters. Moreover, we have proposed an automatic response time extraction algorithm, that has a detection resolution of 50 ms and that can be easily integrated into HRI or inter-human scenarios, even though we designed it for post-experimental audio analysis.

In our study, the participants took more time to answer the card guessing questions during the rounds when they were expected to be deceitful than during those when they were expected to be honest, result which is in line with previous findings. Even though the resolution of our response time extraction algorithm was of 50 ms, which could definitely be improved, it allowed us to detect the differences in response times correlated to the increased cognitive load associated to deception. On top of that, we also proven that these correlations can be observed even without developing a specific interaction scenario that focused specifically on the maximum increase of the cognitive load of participants when they were deceptive.

These results allowed us to improve our previous work, in which we have found correlations between the veracity of the participants' answers and their head position and orientation, heart rate, and skin conductance. The analysis of a larger variety of different parameters and the integration of several techniques could help us improve the robots' ability to detect deception in in-the-wild interaction scenarios. Moreover, combining the polygraph approach and the cognitive load approach

would also yield better overall results, as each of these techniques has its individual strengths and weaknesses.

7.5 My contribution

The main contributions to this chapter was, besides the design and implementation of the experimental setup (including the card game rules, card layout and question trees), the software used to extract the response times automatically.

Chapter 8

Overview of results

8.1 Introduction

In this chapter, all the results and contributions obtained during this thesis are summarised. They are grouped into 4 main categories:

- physiological manifestations correlated to deception
- differences between physiological manifestations in HRI and inter-human interactions
- post-experimental questionnaire evaluation
- publications

In the first two sections, that sum up the differences in physiological manifestations associated to deception as well as to the nature of the interlocutor, we group the results based on the studied physiological manifestations. Then, in the next section, the results of the post-experimental questionnaires are discussed. Moreover, we also present how the participants evaluated our experimental design, what was their perception of the stress level during the experiments, and lastly, to what extent they considered that their interlocutor was able to detect deception. Finally, we resume the list of the accepted scientific publications and conferences we have attended based on the work presented in this thesis.

8.2 Physiological manifestations associated to deception

The results presented in this section have been obtained from two different experimental setups, detailed in chapter 4 and chapter 5: the mock-up crime and interrogation setup and the card guessing game setup. The measurement techniques and devices have varied as well between the two setups, but some of the parameters have been monitored in both of them.

We have used a series of abbreviations to define the experimental conditions in which each result was obtained. The results are differentiated based on the nature of the interaction partner (**RP** = robot partner, **HP** = human partner, **BP** = both robot and human partners) as well as based on the experimental setup that was used to observe these results (**MCI** = mock-up crime and interrogation, **CGG** = card guessing game). In all the tables shown in this section, as it was previously the case, the cells coloured in light green indicate which of the two median values is higher (that associated to truth or that associated to deception).

8.2.1 Heart rate

The following table shows the summary of the results related to the evolution of the participants' heart rate relative to the reference interval, as measured in our experiments, the numeric values corresponding to the **median value** of each measure:

TABLE 8.1: Results summary: heart rate

Measure	$\tilde{x}$: lie	$\tilde{x}$: truth	Conditions	Analysis
median (est. val.)	$-0.24BPM$	$+12BPM$	MCI, RP, female participants	higher when telling the truth
median (est. val.)	$-11.1BPM$	$+0.6BPM$	MCI, RP, male participants	higher when telling the truth
median (est. val.)	$-1.26BPM$	$+10.02BPM$	MCI, RP, participants with high neuroticism	higher when telling the truth
lin. interp. slope	$-3.39 * 10^{-2}$	$-6.66 * 10^{-2}$	CGG, RP, all participants	decreasing at a lower rate when lying
lin. interp. slope	$-6.65 * 10^{-3}$	$2.74 * 10^{-2}$	CGG, HP, all participants	decreasing when lying, increasing when honest

On one hand, these results show that, in our mock-up crime and interrogation experiment, the median estimated heart rate of several experimental subgroups was lower when lying than when telling the truth, in an interrogation scenario undergone by a **robot**. On the other hand, in our card guessing game experiment, when the participants played the game with a **robot partner** we have observed that the participants heart rate decreased both when lying and when telling the truth while answering questions, but that it **decreased at a lower rate** during the rounds when they were lying than during those when they were telling the truth. However, when the participants played the same game with a **human partner**, their heart rate **increased** during the rounds when they were honest and decreased during those when they were lying.

These results indicate that the evolution of the subjects' heart rate can be correlated to their honesty, but that its evolution is different depending on the nature of the interaction partner (robot or human).

8.2.2 Skin conductance

The following table shows the summary of the results related to the evolution of the participants' skin conductance relative to the reference interval, as measured in our experiments, the numeric values corresponding to the **median value** of each measure:

TABLE 8.2: Results summary: skin conductance

Measure	$\tilde{x}$: lie	$\tilde{x}$: truth	Conditions	Analysis
lin. interp. slope	0.00	$1.93 * 10^{-4}$	CGG, RP, all participants	increasing when honest

In our card guessing game experiment, the participants' skin conductance increased during the rounds when they were telling the truth, while staying constant during the rounds when they were deceptive. The increase in the skin conductance implies a decrease of the skin resistance, which is usually correlated to an increased amount of sweating. These results are similar to those obtained by Reid and Inbau [9] in inter-human interactions, as they have observed that the subjects' skin conductance increases more after giving a dishonest answer compared to giving an honest answer to a question.

8.2.3 Head position and orientation

The following table shows the summary of the results related to the evolution of the participants' head position and orientation relative to the reference interval, as measured in our experiments, the numeric values corresponding to the **median value** of each measure:

TABLE 8.3: Results summary: head position and orientation

Parameter	Measure	$\tilde{x}$: lie	$\tilde{x}$: truth	Conditions	Analysis
head height	lin. interp. slope	$\mathbf{5.12 * 10^{-2}}$	$\mathbf{2.67 * 10^{-2}}$	CGG, BP, all participants	increasing at a higher rate when lying
head height	lin. interp. slope	$\mathbf{7.26 * 10^{-2}}$	$\mathbf{1.26 * 10^{-2}}$	CGG, HP, all participants	increasing at a higher rate when lying
head height	lin. interp. slope	$\mathbf{6.86 * 10^{-2}}$	$\mathbf{1.81 * 10^{-2}}$	CGG, BP, participants with high EPQ-L	increasing at a higher rate when lying
head y	average	$\mathbf{-7.75 * 10^{-1}}$	$\mathbf{3.78 * 10^{-1}}$	CGG, RP, all participants	higher position when lying *(inv. y axis)*
head roll	lin. interp. slope	$\mathbf{-1.61 * 10^{-3}}$	$\mathbf{-2.38 * 10^{-3}}$	CGG, BP, participants with low EPQ-L	clockwise increase, at a lower rate when lying

All these results have been obtained in our card guessing game experimental setup. The head height of the participants, which is inversely proportional to the interaction distance, increased more during the rounds when they were deceitful than during the rounds when they were honest, which implicitly means that the **interaction distance decreased more** towards the end of the rounds when they were deceptive than towards the end of the rounds when they were honest. This behaviour was noticed when playing the game with either of the game partners and on the entire experimental group, but also on the experimental subgroup of participants with a high EPQ-L value, as well as on the experiments undergone only with a human partner.

Moreover, the participants' average head vertical (y) position was **higher** during the rounds when they were deceptive (due to the inverted y axis measurements) than during the rounds when they were honest, for the data acquired from the games played with a robot game partner. Lastly, the head roll of the participants with a **low EPQ-L value** increased in a **clockwise direction** both during the rounds when they were honest and during those when they were deceptive, but **at a lower rate when lying**.

8.2.4 Response times to questions

The following table shows the summary of the results related to the evolution of the participants' response time to questions, as measured in our experiments, the numeric values corresponding to the **median value** of each measure:

TABLE 8.4: Results summary: response times

Measure	$\tilde{x}$: lie	$\tilde{x}$: truth	Conditions	Analysis
extracted value	$\mathbf{550}ms$	$\mathbf{500}ms$	CGG, BP, all participants	higher when lying

In our card guessing game experiment, the median value of the participants' measured response time to questions was **higher during the rounds when they were deceptive** than during the rounds when they were honest. Our results are similar to those previously obtained by researchers in both inter-human interaction scenarios [20] [21] [53] as well as HRI scenarios [41] [42].

8.3 HRI vs. inter-human interactions

The results presented in this section have been obtained using the data collected in our **card guessing game** experiment, detailed in chapter 6. In all the tables presented in this section, as it was previously the case, the cells coloured in light green indicate which of the two median values is higher (the value associated to the games played with a robot partner or the value associated to the games played with a human partner).

8.3.1 Skin conductance

The following table shows the summary of the results related to the evolution of the participants' skin conductance relative to the reference interval, as measured in our experiments, the numeric values corresponding to the **median value** of each measure:

TABLE 8.5: Results summary: skin conductance

Measure	$\tilde{x}$: human	$\tilde{x}$: robot	Analysis
average	$\mathbf{-1.15 * 10^{-2}}$	$\mathbf{-1.34 * 10^{-2}}$	higher value when playing the game with a human
lin. interp. slope	$\mathbf{7.51 * 10^{-4}}$	$\mathbf{2.20 * 10^{-5}}$	value increasing more when playing the game with a human

The participants' skin conductance was **higher** and also **increased at a higher rate** during the games played with a **human partner** compared to those played with a robot game partner. The increase in the skin conductance implies a decrease of the skin resistance, which is usually correlated to an increased amount of sweating.

8.3.2 Eye openness

The following table shows the summary of the results related to the evolution of the participants' eye openness relative to the reference interval, as measured in our experiments, the numeric values corresponding to the **median value** of each measure:

TABLE 8.6: Results summary: eye openness

Measure	$\tilde{x}$: human	$\tilde{x}$: robot	Analysis
average	$\mathbf{-1.15 * 10^{-2}}$	$\mathbf{-1.34 * 10^{-2}}$	eyes more open when playing the game with a human
std. dev.	$\mathbf{7.80 * 10^{-3}}$	$\mathbf{2.25 * 10^{-3}}$	more eyelid movements when playing the game with a human

The participants' eye openness was **higher** and also had a **higher standard deviation** during the games played with a **human partner** compared to those played with a robot game partner. The higher standard deviation of the eye openness implies that the participants **blinked more during the games played with a human** game partner than during those played with a robot game partner.

8.3.3 Head position and orientation

The following table shows the summary of the results related to the evolution of the participants' skin conductance relative to the reference interval, as measured in our experiments, the numeric values corresponding to the **median value** of each measure:

TABLE 8.7: Results summary: head position and orientation

Parameter	Measure	$\tilde{x}$: human	$\tilde{x}$: robot	Analysis
Head x	std. dev.	$\mathbf{2.71 * 10^{-1}}$	$\mathbf{4.90 * 10^{-1}}$	more variations when playing the game with a robot
head y	std. dev.	$\mathbf{1.32 * 10^{-1}}$	$\mathbf{2.27 * 10^{-1}}$	more variations when playing the game with a robot
head width	std. dev.	$\mathbf{5.11 * 10^{-1}}$	$\mathbf{4.02 * 10^{-2}}$	more variations when playing the game with a human
head height	std. dev.	$\mathbf{4.97 * 10^{-1}}$	$\mathbf{5.06 * 10^{-2}}$	more variations when playing the game with a human
head roll	average	$\mathbf{3.05 * 10^{-2}}$	$\mathbf{1.12 * 10^{-2}}$	higher value when playing the game with a human
head pan	std. dev.	$\mathbf{-2.48 * 10^{-2}}$	$\mathbf{3.34 * 10^{-2}}$	more variations when playing the game with a robot
head tilt	std. dev.	$\mathbf{-8.01 * 10^{-3}}$	$\mathbf{4.17 * 10^{-2}}$	more variations when playing the game with a robot

The **standard deviations** of the participants' head **x** and **y** coordinates, as well as their **pan** and **tilt** angles, were **higher** when playing the game with a **robot partner** than when playing the game with a human partner. This means that the participants moved their heads up and down more and also rotated it more during the games played with a robot game partner. On the other hand, the standard deviations of the participants **height** and **width** were **lower** when playing the game with a **robot partner**, which means that the participants' **interaction distance varied less** during the games played with a robot partner. A more constant interaction distance is a usually associated with a more controlled upper body posture when sitting down at the table.

Lastly, the participants' **average head roll** was higher, which means that the participants' head was rotated more in a clockwise direction, during the games played with a human partner than during those played with a robot partner.

8.4 Post-experimental questionnaire evaluation

We have asked the participants to our experiments to fill in various post-experimental questionnaires. Their purpose was to evaluate the participants' perception of various aspects of our experimental setups, as well as of their own emotional state. The results of these questionnaires have already been detailed in the previous chapters, but we sum up in this section three main aspects of interest for us.

Experimental design evaluation

One of the aspects we wanted to evaluate using post-experimental questionnaires was how motivated the participants were to win their reward, or, from a different perspective, how enticed they were to deceive their interaction partner.

In both our experimental setups, the participants' desire to win the assigned rewards was well above average, with better reported values in the case of the second experiment. In our first experimental setup, the average desirability of the chocolate bar was of **5.375**, while in our second experimental setup the desirability of the candy was of **5.14**, while the desirability of the 50 € gift card was of **5.80**. All values are expressed on a scale from 1 to 7.

Moreover, the participants liked playing our card guessing game, with a reported average value of **5.74**, and also wanted it to play it again, with an average value of **4.80**.

The perception of stress

Another factor that we were interested in evaluating was the participants' perception of stress during our experiments. More specifically, we wanted to make sure that the stress level perceived by the participants was as low as possible. In the first

experiment, the participants have reported an average stress level of **1.667** during the mock-up crime phase and of **3.792** during the interrogation phase, both being under average. In our second experimental setup, the card guessing game, the reported stress level was even lower than during the interrogation phase of the mock-up crime experiment, with a reported average value of **3.68** during the rounds when the participants were supposed to be deceitful and of **2.56** during those when they were supposed to be honest. Overall, the reported stress levels were low.

On a similar note, we evaluated to what extent the participants considered difficult or frustrating to answer the questions that were asked during our card guessing game experiment. In both cases, the reported values were low. The average values of the perceived difficulty of **2.48** (honest rounds) and **3.00** (deceitful rounds), while the average values of the perceived frustration were of **1.98** and **2.56**, respectively.

The perception of the ability to detect deception

Lastly, we were interested in evaluating to what extent the participants believed their interaction partner was able to detect deception during the experiment. In the case of our first experimental setup, the participants had a slightly below average confidence in the robots' ability to detect deception, with an average reported value of **3.708**, while also being slightly above average convinced that a human interrogator would have been more capable of detecting when they were deceitful, with an average value of **4.500**.

In our second experiment, the participants had an even lower confidence in their interaction partner's ability to detect deception, regardless of their nature, human or robot. More specifically, the average reported values were of **3.20** for the human game partner and of **2.80** for the robot game partner. All in all, if we take into account the results of both our experiments, participants tend to credit the robot's ability to detect deception less than a human's ability, while considering that both their abilities to detect deception are below average.

8.5 Publications

Published

- ICSR 2019, Madrid, Spain: *The impact of a robot game partner when studying deception during a card game* (David-Octavian Iacob, Adriana Tapus)
- RO-MAN 2019, New Delhi, India: *Detecting deception in HRI using minimally-invasive and noninvasive techniques* (David-Octavian Iacob, Adriana Tapus)
- RO-MAN 2018, Nanjing, China: *First Attempts in Deception Detection in HRI by using Thermal and RGB-D cameras* (David-Octavian Iacob, Adriana Tapus)

Chapter 9

Conclusions

9.1 Introduction

The main goal of this thesis was to improve the robots' ability to detect deception, therefore making them more capable of interacting naturally and efficiently with humans. In the previous chapters, we studied the state of the art in deception detection in inter-human and human-robot interactions, we established our experimental design guidelines, then detailed our attempts to accomplish our goal and ended by summarising the list of results we obtained, as well as of the scientific works we published. This chapter concludes this thesis by analysing the significance of our results, as well as the perspectives that our research work has opened.

9.2 The significance of our results

In chapter 8, we compiled a summary of the results we obtained throughout our work. Without going again through the entire list of parameters we analysed and the correlations we found, it is important to understand the relevance of these results and how they contribute to achieving our main goal. On one hand, we found a series of correlations between various physiological or behavioural manifestations and the veracity of the answers given by the participants to our experiments. Some of these correlations were found in both inter-human and human-robot interaction scenarios, while others only in one of the two experimental conditions. Some of the results obtained when participants interacted with robots confirmed the previous findings and theories established in inter-human interaction scenarios, while others did not.

On the other hand, we also directly compared the physiological and behavioural manifestations exhibited by the participants to our experiments when interacting with a robot and a human interlocutor. Our findings show that there is indeed a correlation between the nature of the interlocutor - human or robot - and some of the manifestations exhibited by the participants, at least in the case of the interaction scenarios we studied, where the latter were enticed to lie. The parameters we analysed were the same in both cases - when studying the manifestations associated to deception as well when studying the impact of the interaction partner.

There are two major conclusions we can draw based on these results. Firstly, we showed that humans who deceive their robot interlocutors do indeed exhibit a series of manifestations that can be analysed and identified by a robotic system. Moreover, the tools and methods we used to identify these manifestations do not hinder in any significant way the quality of the interaction, thus rendering them usable in a large variety of HRI scenarios. Secondly, these cues associated to deception that we identified in our HRI scenarios are not always the same as those identified in

inter-human interaction scenarios, neither by us nor by other researchers who have studied the mechanisms of deception in the past.

We succeeded in understanding some of the mechanisms associated to deception in HRI, in finding appropriate methods that allow robots to detect them in real time, as well as in showing that the nature of the interlocutor has an impact on the cues associated to deception. The end result of our work is that robots are one step closer to being able to detect if their human interlocutors are deceitful and, implicitly, one step closer to better understanding their human interlocutors.

This would only help improving the quality of the human-robot interactions, as well as the utility of social and socially assistive robots, that will be more capable of performing the tasks they are required to and more capable of improving the quality of life of their users.

9.3 Perspectives

Our work focused mainly on understanding the mechanisms of deception in HRI and on proposing appropriate devices and methods that allow robots to detect these mechanisms. Therefore, the results and conclusions of our work need to be taken one step further in order to be translated into the development of social robots that are able to reliably detect when their human interlocutors deceive them.

This extra step is represented, in our opinion, by the development of a classification algorithm, using appropriate machine learning techniques, that is able to decide, based on the real-time analysis of various manifestations, if an interlocutor is honest or not at a given point during the interaction. An attempt to develop such a system was already made by Gonzalez-Billandon, Aroyo et al. [42], who used measurements of the pupil dilation, saccades, and response times to questions and train a classifier using the data samples they have acquired throughout their experiments. They have managed to obtain a lie detection accuracy of up to 73% and a precision of up to 74%, thus proving that it is indeed possible to train a classifier that is able to detect deceitful behaviour in HRI with an accuracy and precision similar to those of some state of the art deception detection systems.

Using a larger number of features, represented by the various manifestations we have studied, would improve such a classifier's accuracy and precision, especially in the interaction scenarios where the robot is prone to fail to acquire reliable measures from all channels simultaneously. Moreover, a larger variety of physiological or behavioural parameters that are analysed also implies that the system can go one step further than its *general* prediction model, trained on a large number of samples acquired from various users, by learning which of the cues associated to deception are more reliable when interacting with **one particular user**. This objective can be achieved by means of long-term, unsupervised, or semi-supervised learning, therefore modelling a similar process to that undergone by humans on a daily basis, as they spend more time and get familiar with a number of people who are part of their lives - close friends, family or coworkers. Robots in general and socially assistive robots in particular would benefit from such an approach, which would be more realistic and appropriate than the development of a system that does not adapt in any way to its users.

Lastly, besides developing a classifier that would ideally allow a robot to decide whether their interlocutor is honest or not, henceforth improving its ability to have a better understanding of humans and better interaction capabilities, it is important to also significantly improve the physiological and behavioural measurement analysis

capabilities of robots. One of the major source of setbacks in our research work was the lack of readily-available, standardised, noninvasive or minimally-invasive measurement devices and techniques, that would allow robots to evaluate and quantify their interlocutor's manifestations accurately and robustly. This represents, from our perspective, the weakest link in the scientific and technological chain required to improve the ability of robots to detect deception. Much as deception detection always was a difficult task, with even the most advanced systems being far from perfect, the lack of proper measurement devices and techniques, adapted to the specifics of *in-the-wild* HRI makes it even more gruelling.

Considering the directions in which we believe this topic should continue evolving, as well as our contribution to it, we strongly believe that it is just a matter of time until robots will eventually be able to detect deception at least as accurately as trained or experienced humans. We also believe that this ability to detect deception represents one of the many bricks upon which the social interaction abilities of robots are built, and that our work has contributed to the global goal of social robotics, which is to improve the quality of life of human users by means of social interactions.

Academic CV

Publications

Published

- ICSR 2019, Madrid, Spain: *The impact of a robot game partner when studying deception during a card game* (David-Octavian Iacob, Adriana Tapus)
- RO-MAN 2019, New Delhi, India: *Detecting deception in HRI using minimally-invasive and noninvasive techniques* (David-Octavian Iacob, Adriana Tapus)
- RO-MAN 2018, Nanjing, China: *First Attempts in Deception Detection in HRI by using Thermal and RGB-D cameras* (David-Octavian Iacob, Adriana Tapus)
- RO-MAN 2016, New York, USA: *Joint Attention using Human-Robot Interaction: Impact of sensory preferences of children with autism* (Pauline Chevalier, Jean-Claude Martin, Brice Isableu, Christophe Bazile, David-Octavian Iacob, Adriana Tapus)

Reviews

- THRI 2019
- Frontiers in Robotics and AI 2019
- ECMR 2019
- ECMR 2017

Teaching

I have been in charge of laboratory assignments at ENSTA Paris for the following courses:

- **2016-2017**
  - **IN102** *(Langage C / C language)* - Prof. François Pessaux
  - **IN103** *(Algorithmique (langage C) / Algorithmics (C language)* - Prof. François Pessaux
- **2017-2018**
  - **IN102** *(Langage C / C language)* - Prof. François Pessaux
  - **IN103** *(Algorithmique (en C) / Algorithmics (in C)* - Prof. François Pessaux
- **2018-2019**

  - **IN101** *(Algorithmique et programmation (en Python) / Algorithmics and programming (in C)* - Prof. François Pessaux
  - **IN102** *(Système et programmation en C / C System and programming)* - Prof. Goran Frehse
  - **IN103** *(Système et programmation en C (partie 2) / C System and programming (second part))* - Prof. Julien Alexandre dit Sandretto
  - **ROB311** *(Apprentissage pour la robotique / Machine learning for robotics)* - Prof. Adriana Tapus

- **2019-2020**

  - **ROB311** *(Apprentissage pour la robotique / Machine learning for robotics)* - Prof. Adriana Tapus

Other activities

- student representative in the Council of the INTERFACES Doctoral School (September 2017 - September 2019)
- student representative in the Research Council of ENSTA Paris (September 2017 - September 2019)

Prizes

- winner of the 2017 edition of the *Défi 6 minutes (6 minutes challenge)* contest organized by the INTERFACES Doctoral School

References

[1] J. R. Stroop. "Studies of interference in serial verbal reactions." In: *Journal of Experimental Psychology* 18.6 (1935), pp. 643–662. DOI: 10.1037/h0054651. URL: https://doi.org/10.1037/h0054651.

[2] Maja J Matarić et al. "Socially assistive robotics for post-stroke rehabilitation". In: *Journal of NeuroEngineering and Rehabilitation* 4.1 (Feb. 2007). DOI: 10.1186/1743-0003-4-5. URL: https://doi.org/10.1186/1743-0003-4-5.

[3] Adriana Tapus, Mataric Maja, and Brian Scassellatti. "The Grand Challenges in Socially Assistive Robotics". In: *IEEE Robotics and Automation Magazine* 14.1 (2007), N/A. URL: https://hal.archives-ouvertes.fr/hal-00770113.

[4] Mark Frank, Melissa A. Menasco, and Maureen O'Sullivan. "Human Behavior and Deception Detection". In: Nov. 2008. ISBN: 9780470087923. DOI: 10.1002/9780470087923.hhs299.

[5] Charles F. Bond and Bella M. DePaulo. "Accuracy of Deception Judgments". In: *Personality and Social Psychology Review* 10.3 (Aug. 2006), pp. 214–234. DOI: 10.1207/s15327957pspr1003_2. URL: https://doi.org/10.1207/s15327957pspr1003_2.

[6] Samantha Mann, Aldert Vrij, and Ray Bull. "Detecting True Lies: Police Officers' Ability to Detect Suspects' Lies." In: *Journal of Applied Psychology* 89.1 (2004), pp. 137–149. DOI: 10.1037/0021-9010.89.1.137. URL: https://doi.org/10.1037/0021-9010.89.1.137.

[7] Pär Anders Granhag, Aldert Vrij, and Bruno Verschuere. *Detecting Deception: Current Challenges and Cognitive Approaches (Wiley Series in Psychology of Crime, Policing and Law)*. Wiley-Blackwell, 2014. ISBN: 9781118509753. URL: https://www.amazon.com/Detecting-Deception-Challenges-Approaches-Psychology-ebook/dp/B00P3VSC24?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=B00P3VSC24.

[8] John Synnott, David Dietzel, and Maria Ioannou. "A review of the polygraph: history, methodology and current status". In: *Crime Psychology Review* 1.1 (2015), pp. 59–83. DOI: 10.1080/23744006.2015.1060080. eprint: https://doi.org/10.1080/23744006.2015.1060080. URL: https://doi.org/10.1080/23744006.2015.1060080.

[9] John E. Reid and Fred E. Inbau. *Truth and Deception: The Polygraph (Lie-Detector Technique)*. Williams & Wilkins, 1977. ISBN: 0683072242. URL: https://www.amazon.com/Truth-Deception-Polygraph-Lie-Detector-Technique/dp/0683072242?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0683072242.

[10] John E. Reid. "A Revised Questioning Technique in Lie-Detection Tests". In: *Journal of Criminal Law and Criminology (1931-1951)* 37.6 (Mar. 1947), p. 542. DOI: 10.2307/1138979. URL: https://doi.org/10.2307/1138979.

[11] Frank S. Horvath and John E. Reid. "The Reliability of Polygraph Examiner Diagnosis of Truth and Deception". In: *The Journal of Criminal Law, Criminology, and Police Science* 62.2 (June 1971), p. 276. DOI: 10.2307/1141892. URL: https://doi.org/10.2307/1141892.

[12] Andrea Gaggioli. "Beyond the Truth Machine: Emerging Technologies for Lie Detection". In: *Cyberpsychology, Behavior, and Social Networking* 21.2 (Feb. 2018), pp. 144–144. DOI: 10.1089/cyber.2018.29102.csi. URL: https://doi.org/10.1089/cyber.2018.29102.csi.

[13] W. G. Iacono and D. T. Lykken. "The validity of the lie detector: Two surveys of scientific opinion." In: *Journal of Applied Psychology* 82.3 (1997), pp. 426–433. DOI: 10.1037/0021-9010.82.3.426. URL: https://doi.org/10.1037/0021-9010.82.3.426.

[14] Tatia M.C. Lee et al. "Lie detection by functional magnetic resonance imaging". In: *Human Brain Mapping* 15.3 (Jan. 2002), pp. 157–164. DOI: 10.1002/hbm.10020. URL: https://doi.org/10.1002/hbm.10020.

[15] I. Pavlidis and J. Levine. "Thermal facial screening for deception detection". In: *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society] [Engineering in Medicine and Biology*. IEEE. DOI: 10.1109/iembs.2002.1106317. URL: https://doi.org/10.1109/iembs.2002.1106317.

[16] Sarun Sumriddetchkajorn and Armote Somboonkaew. "Thermal analyzer enables improved lie detection in criminal-suspect interrogations". In: *SPIE Newsroom* (Feb. 2011). DOI: 10.1117/2.1201101.003452. URL: https://doi.org/10.1117/2.1201101.003452.

[17] Sharon Leal and Aldert Vrij. "Blinking During and After Lying". In: *Journal of Nonverbal Behavior* 32.4 (July 2008), pp. 187–194. DOI: 10.1007/s10919-008-0051-0. URL: https://doi.org/10.1007/s10919-008-0051-0.

[18] Nisha Bhaskaran et al. "Lie to Me: Deceit detection via online behavioral learning". In: *Face and Gesture 2011*. IEEE, Mar. 2011. DOI: 10.1109/fg.2011.5771407. URL: https://doi.org/10.1109/fg.2011.5771407.

[19] Fatih Veysel Nurçin et al. "Lie detection on pupil size by back propagation neural network". In: *Procedia Computer Science* 120 (2017), pp. 417–421. DOI: 10.1016/j.procs.2017.11.258. URL: https://doi.org/10.1016/j.procs.2017.11.258.

[20] Miron Zuckerman, Bella M. DePaulo, and Robert Rosenthal. "Verbal and Nonverbal Communication of Deception". In: *Advances in Experimental Social Psychology*. Elsevier, 1981, pp. 1–59. DOI: 10.1016/s0065-2601(08)60369-x. URL: https://doi.org/10.1016/s0065-2601(08)60369-x.

[21] Aldert Vrij et al. "A cognitive load approach to lie detection". In: *Journal of Investigative Psychology and Offender Profiling* 5.1-2 (2008), pp. 39–43. DOI: 10.1002/jip.82. URL: https://doi.org/10.1002/jip.82.

[22] Aldert Vrij et al. "Outsmarting the Liars: Toward a Cognitive Lie Detection Approach". In: *Current Directions in Psychological Science* 20.1 (Feb. 2011), pp. 28–32. DOI: 10.1177/0963721410391245. URL: https://doi.org/10.1177/0963721410391245.

[23] Aldert Vrij et al. "Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order." In: *Law and Human Behavior* 32.3 (2008), pp. 253–265. DOI: 10.1007/s10979-007-9103-y. URL: https://doi.org/10.1007/s10979-007-9103-y.

[24] Jeffrey J. Walczyk et al. "Advancing Lie Detection by Inducing Cognitive Load on Liars: A Review of Relevant Theories and Techniques Guided by Lessons from Polygraph-Based Approaches". In: *Frontiers in Psychology* 4 (2013). DOI: 10.3389/fpsyg.2013.00014. URL: https://doi.org/10.3389/fpsyg.2013.00014.

[25] *Fundamentals of Anatomy and Physiology: For Nursing and Healthcare Students*. Wiley-Blackwell, 2016, pp. 426–429. ISBN: 1119055520. URL: https://www.amazon.com/Fundamentals-Anatomy-Physiology-Healthcare-Students/dp/1119055520?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=1119055520.

[26] John A. Podlesny and David C. Raskin. "Effectiveness of Techniques and Physiological Measures in the Detection of Deception". In: *Psychophysiology* 15.4 (July 1978), pp. 344–359. DOI: 10.1111/j.1469-8986.1978.tb01391.x. URL: https://doi.org/10.1111/j.1469-8986.1978.tb01391.x.

[27] David C. Raskin and Robert D. Hare. "Psychopathy and Detection of Deception In a Prison Population". In: *Psychophysiology* 15.2 (Mar. 1978), pp. 126–136. DOI: 10.1111/j.1469-8986.1978.tb01348.x. URL: https://doi.org/10.1111/j.1469-8986.1978.tb01348.x.

[28] M. T. Bradley and Michel Pierre Janisse. "Accuracy Demonstrations, Threat, and the Detection of Deception: Cardiovascular, Electrodermal, and Pupillary Measures". In: *Psychophysiology* 18.3 (May 1981), pp. 307–315. DOI: 10.1111/j.1469-8986.1981.tb03040.x. URL: https://doi.org/10.1111/j.1469-8986.1981.tb03040.x.

[29] William M. Marston. "Systolic blood pressure symptoms of deception." In: *Journal of Experimental Psychology* 2.2 (1917), pp. 117–163. DOI: 10.1037/h0073583. URL: https://doi.org/10.1037/h0073583.

[30] John E. Reid. "Simulated Blood Pressure Responses in Lie-Detector Tests and a Method for Their Detection". In: *Journal of Criminal Law and Criminology (1931-1951)* 36.3 (Sept. 1945), p. 201. DOI: 10.2307/1138525. URL: https://doi.org/10.2307/1138525.

[31] Robert J. Cutrow et al. "The Objective Use of Multiple Physiological Indices in the Detection of Deception". In: *Psychophysiology* 9.6 (Nov. 1972), pp. 578–588. DOI: 10.1111/j.1469-8986.1972.tb00767.x. URL: https://doi.org/10.1111/j.1469-8986.1972.tb00767.x.

[32] I. Pavlidis and J. Levine. "Monitoring of periorbital blood flow rate through thermal image analysis and its application to polygraph testing". In: *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. DOI: 10.1109/iembs.2001.1017374. URL: https://doi.org/10.1109/iembs.2001.1017374.

[33] Dean A. Pollina et al. "Facial Skin Surface Temperature Changes During a "Concealed Information" Test". In: *Annals of Biomedical Engineering* 34.7 (June 2006), pp. 1182–1189. DOI: 10.1007/s10439-006-9143-3. URL: https://doi.org/10.1007/s10439-006-9143-3.

[34] Bashar A. Rajoub and Reyer Zwiggelaar. "Thermal Facial Analysis for Deception Detection". In: *IEEE Transactions on Information Forensics and Security* 9.6 (June 2014), pp. 1015–1023. DOI: 10.1109/tifs.2014.2317309. URL: https://doi.org/10.1109/tifs.2014.2317309.

[35] Daniel Kahneman. *Attention and effort (Prentice-Hall series in experimental psychology)*. Prentice-Hall, 1973. ISBN: 0130505188. URL: https://www.amazon.com/Attention-effort-Prentice-Hall-experimental-psychology/dp/0130505188?SubscriptionId=AKIAIOBINVZYXZQZ2U3A&tag=chimbori05-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0130505188.

[36] Ira Heilveil. "Deception and pupil size". In: *Journal of Clinical Psychology* 32.3 (July 1976), pp. 675–676. DOI: 10.1002/1097-4679(197607)32:3<675::aid-jclp2270320340>3.0.co;2-a. URL: https://doi.org/10.1002/1097-4679(197607)32:3%3C675::aid-jclp2270320340%3E3.0.co;2-a.

[37] M Bradley and Michel P. Janisse. "Pupil size and lie detection: The effect of certainty on detection". In: *Psychology: A Journal of Human Behavior* 16 (Jan. 1970), pp. 33–39.

[38] Michel Pierre Janisse and M Bradley. "Deception, information and the pupillary response". In: *Perceptual and Motor Skills* 50 (June 1980), pp. 748–750. DOI: 10.2466/pms.1980.50.3.748.

[39] Robert Lubow and Ofer Fein. "Pupillary Size in Response to a Visual Guilty Knowledge Test: New Technique for the Detection of Deception". In: *Journal of Experimental Psychology: Applied* 2 (June 1996), pp. 164–177. DOI: 10.1037/1076-898X.2.2.164.

[40] Travis L. Seymour, Christopher A. Baker, and Joshua T. Gaunt. "Combining Blink, Pupil, and Response Time Measures in a Concealed Knowledge Test". In: *Frontiers in Psychology* 3 (2013). DOI: 10.3389/fpsyg.2012.00614. URL: https://doi.org/10.3389/fpsyg.2012.00614.

[41] A.M. Aroyo et al. "Can a Humanoid Robot Spot a Liar?" In: *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*. IEEE, Nov. 2018. DOI: 10.1109/humanoids.2018.8624992. URL: https://doi.org/10.1109/humanoids.2018.8624992.

[42] Jonas Gonzalez-Billandon et al. "Can a Robot Catch You Lying? A Machine Learning System to Detect Lies During Interactions". In: *Frontiers in Robotics and AI* 6 (July 2019). DOI: 10.3389/frobt.2019.00064. URL: https://doi.org/10.3389/frobt.2019.00064.

[43] D.D. Langleben et al. "Brain Activity during Simulated Deception: An Event-Related Functional Magnetic Resonance Study". In: *NeuroImage* 15.3 (Mar. 2002), pp. 727–732. DOI: 10.1006/nimg.2001.1003. URL: https://doi.org/10.1006/nimg.2001.1003.

[44] F. Andrew Kozel et al. "A Pilot Study of Functional Magnetic Resonance Imaging Brain Correlates of Deception in Healthy Young Men". In: *The Journal of Neuropsychiatry and Clinical Neurosciences* 16.3 (Aug. 2004), pp. 295–305. DOI: 10.1176/jnp.16.3.295. URL: https://doi.org/10.1176/jnp.16.3.295.

[45] C. Davatzikos et al. "Classifying spatial patterns of brain activity with machine learning methods: Application to lie detection". In: *NeuroImage* 28.3 (Nov. 2005), pp. 663–668. DOI: 10.1016/j.neuroimage.2005.08.009. URL: https://doi.org/10.1016/j.neuroimage.2005.08.009.

[46] Bella M. DePaulo et al. "Cues to deception." In: *Psychological Bulletin* 129.1 (2003), pp. 74–118. DOI: 10.1037/0033-2909.129.1.74. URL: https://doi.org/10.1037/0033-2909.129.1.74.

[47] Shan Lu et al. "Blob Analysis of the Head and Hands: A Method for Deception Detection". In: *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*. IEEE, 2005. DOI: 10.1109/hicss.2005.122. URL: https://doi.org/10.1109/hicss.2005.122.

[48] T.O. Meservy et al. "Deception Detection through Automatic, Unobtrusive Analysis of Nonverbal Behavior". In: *IEEE Intelligent Systems* 20.5 (Sept. 2005), pp. 36–43. DOI: 10.1109/mis.2005.85. URL: https://doi.org/10.1109/mis.2005.85.

[49] J.K. Burgoon et al. "Detecting Concealment of Intent in Transportation Screening: A Proof of Concept". In: *IEEE Transactions on Intelligent Transportation Systems* 10.1 (Mar. 2009), pp. 103–112. DOI: 10.1109/tits.2008.2011700. URL: https://doi.org/10.1109/tits.2008.2011700.

[50] Jia E. Loy, Hannah Rohde, and Martin Corley. "Cues to Lying May be Deceptive: Speaker and Listener Behaviour in an Interactive Game of Deception". In: *Journal of Cognition* 1.1 (2018). DOI: 10.5334/joc.46. URL: https://doi.org/10.5334/joc.46.

[51] Aldert Vrij and Samantha Mann. "Telling and detecting lies in a high-stake situation: the case of a convicted murderer". In: *Applied Cognitive Psychology* 15.2 (2000), pp. 187–203. DOI: 10.1002/1099-0720(200103/04)15:2<187::aid-acp696>3.0.co;2-a. URL: https://doi.org/10.1002/1099-0720(200103/04)15:2%3C187::aid-acp696%3E3.0.co;2-a.

[52] Siegfried L. Sporer and Barbara Schwandt. "Moderators of nonverbal indicators of deception: A meta-analytic synthesis." In: *Psychology, Public Policy, and Law* 13.1 (2007), pp. 1–34. DOI: 10.1037/1076-8971.13.1.1. URL: https://doi.org/10.1037/1076-8971.13.1.1.

[53] Jeffrey J. Walczyk et al. "Cognitive mechanisms underlying lying to questions: response time as a cue to deception". In: *Applied Cognitive Psychology* 17.7 (2003), pp. 755–774. DOI: 10.1002/acp.914. URL: https://doi.org/10.1002/acp.914.

[54] G. C. Drew. "Variations in Reflex Blink-Rate during Visual-Motor Tasks". In: *Quarterly Journal of Experimental Psychology* 3.2 (Apr. 1951), pp. 73–88. DOI: 10.1080/17470215108416776. URL: https://doi.org/10.1080/17470215108416776.

[55] Janice Bagley and Leon Manelis. "Effect of Awareness on an Indicator of Cognitive Load". In: *Perceptual and Motor Skills* 49.2 (Dec. 1979), pp. 591–594. DOI: 10.2466/pms.1979.49.2.591. URL: https://doi.org/10.2466/pms.1979.49.2.591.

[56] Robert Goldstein, Lance O. Bauer, and John A. Stern. "Effect of task difficulty and interstimulus interval on blink parameters". In: *International Journal of Psychophysiology* 13.2 (Sept. 1992), pp. 111–117. DOI: 10.1016/0167-8760(92)90050-l. URL: https://doi.org/10.1016/0167-8760(92)90050-l.

[57] Morris K. Holland and Gerald Tarlow. "Blinking and Mental Load". In: *Psychological Reports* 31.1 (Aug. 1972), pp. 119–127. DOI: 10.2466/pr0.1972.31.1.119. URL: https://doi.org/10.2466/pr0.1972.31.1.119.

[58] John A. Stern, Larry C. Walrath, and Robert Goldstein. "The Endogenous Eyeblink". In: *Psychophysiology* 21.1 (Jan. 1984), pp. 22–33. DOI: 10.1111/j.1469-8986.1984.tb02312.x. URL: https://doi.org/10.1111/j.1469-8986.1984.tb02312.x.

[59] Aldert Vrij, Pär Anders Granhag, and Stephen Porter. "Pitfalls and Opportunities in Nonverbal and Verbal Lie Detection". In: *Psychological Science in the Public Interest* 11.3 (Dec. 2010), pp. 89–121. DOI: 10.1177/1529100610390861. URL: https://doi.org/10.1177/1529100610390861.

[60] Tadas Baltrusaitis, Peter Robinson, and Louis-Philippe Morency. "Constrained Local Neural Fields for Robust Facial Landmark Detection in the Wild". In: *2013 IEEE International Conference on Computer Vision Workshops*. IEEE, Dec. 2013. DOI: 10.1109/iccvw.2013.54. URL: https://doi.org/10.1109/iccvw.2013.54.

[61] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. "3D Constrained Local Model for Rigid and Non-Rigid Facial Tracking". In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2012.

[62] Davis E. King. "Dlib-ml: A Machine Learning Toolkit". In: *Journal of Machine Learning Research* 10 (2009), pp. 1755–1758.

[63] Chin-An Wang et al. "Arousal Effects on Pupil Size, Heart Rate, and Skin Conductance in an Emotional Face Task". In: *Frontiers in Neurology* 9 (Dec. 2018). DOI: 10.3389/fneur.2018.01029. URL: https://doi.org/10.3389/fneur.2018.01029.

[64] Ibrahim Furkan Ince and Jin Woo Kim. "A 2D eye gaze estimation system with low-resolution webcam images". In: *EURASIP Journal on Advances in Signal Processing* 2011.1 (Aug. 2011). DOI: 10.1186/1687-6180-2011-40. URL: https://doi.org/10.1186/1687-6180-2011-40.

[65] Javier San Agustin et al. "Evaluation of a low-cost open-source gaze tracker". In: *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications - ETRA '10*. ACM Press, 2010. DOI: 10.1145/1743666.1743685. URL: https://doi.org/10.1145/1743666.1743685.

[66] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson. "Remote plethysmographic imaging using ambient light". In: *Optics Express* 16.26 (Dec. 2008), p. 21434. DOI: 10.1364/oe.16.021434. URL: https://doi.org/10.1364/oe.16.021434.

[67] Ming-Zher Poh, Daniel J. McDuff, and Rosalind W. Picard. "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation". In: *Optics Express* 18.10 (May 2010), p. 10762. DOI: 10.1364/oe.18.010762. URL: https://doi.org/10.1364/oe.18.010762.

[68] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. "Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam". In: *IEEE Transactions on Biomedical Engineering* 58.1 (Jan. 2011), pp. 7–11. DOI: 10.1109/tbme.2010.2086456. URL: https://doi.org/10.1109/tbme.2010.2086456.

[69] Denis Laure and Ilya Paramonov. "Improved algorithm for heart rate measurement using mobile phone camera". In: *2013 13th Conference of Open Innovations Association (FRUCT)*. IEEE, Apr. 2013. DOI: 10.23919/fruct.2013.8124232. URL: https://doi.org/10.23919/fruct.2013.8124232.

[70] L Tarassenko et al. "Non-contact video-based vital sign monitoring using ambient light and auto-regressive models". In: *Physiological Measurement* 35.5 (Mar. 2014), pp. 807–831. DOI: 10.1088/0967-3334/35/5/807. URL: https://doi.org/10.1088/0967-3334/35/5/807.

[71] I. Pavlidis, J. Levine, and P. Baukol. "Thermal imaging for anxiety detection". In: *Proceedings IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications (Cat. No.PR00640)*. IEEE Comput. Soc. DOI: 10.1109/cvbvs.2000.855255. URL: https://doi.org/10.1109/cvbvs.2000.855255.

[72] P. Yuen et al. "Emotional & physical stress detection and classification using thermal imaging technique". In: *3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009)*. IET, 2009. DOI: 10.1049/ic.2009.0241. URL: https://doi.org/10.1049/ic.2009.0241.

[73] Roxana Agrigoroaie, Arturo Cruz-Maya, and Adriana Tapus. ""Oh! I am so sorry!": Understanding User Physiological Variation while Spoiling a Game Task". In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Oct. 2018. DOI: 10.1109/iros.2018.8593395. URL: https://doi.org/10.1109/iros.2018.8593395.

[74] Roxana Agrigoroaie and Adriana Tapus. "Cognitive Performance and Physiological Response Analysis". In: *International Journal of Social Robotics* (Mar. 2019). DOI: 10.1007/s12369-019-00532-z. URL: https://doi.org/10.1007/s12369-019-00532-z.

[75] I. Pavlidis et al. "Interacting with human physiology". In: *Computer Vision and Image Understanding* 108.1-2 (Oct. 2007), pp. 150–170. DOI: 10.1016/j.cviu.2006.11.018. URL: https://doi.org/10.1016/j.cviu.2006.11.018.

[76] Marc Garbey et al. "Contact-Free Measurement of Cardiac Pulse Based on the Analysis of Thermal Imagery". In: *IEEE Transactions on Biomedical Engineering* 54.8 (Aug. 2007), pp. 1418–1426. DOI: 10.1109/tbme.2007.891930. URL: https://doi.org/10.1109/tbme.2007.891930.

[77] Ronan Chauvin et al. "Contact-Free Respiration Rate Monitoring Using a Pan–Tilt Thermal Camera for Stationary Bike Telerehabilitation Sessions". In: *IEEE Systems Journal* 10.3 (Sept. 2016), pp. 1046–1055. DOI: 10.1109/jsyst.2014.2336372. URL: https://doi.org/10.1109/jsyst.2014.2336372.

[78] J. Rumiński. "Evaluation of respiration rate and pattern using a portable thermal camera". In: *Proceedings of the 2016 International Conference on Quantitative InfraRed Thermography*. QIRT Council, 2016. DOI: 10.21611/qirt.2016.107. URL: https://doi.org/10.21611/qirt.2016.107.

[79] Fernando González et al. "Smart Multi-Level Tool for Remote Patient Monitoring Based on a Wireless Sensor Network and Mobile Augmented Reality". In: *Sensors* 14.9 (Sept. 2014), pp. 17212–17234. DOI: 10.3390/s140917212. URL: https://doi.org/10.3390/s140917212.

[80] David Giles, Nick Draper, and William Neil. "Validity of the Polar V800 heart rate monitor to measure RR intervals at rest". In: *European Journal of Applied Physiology* 116.3 (Dec. 2015), pp. 563–571. DOI: 10.1007/s00421-015-3303-9. URL: https://doi.org/10.1007/s00421-015-3303-9.

[81] Taisa Daiana da Costa et al. "Breathing Monitoring and Pattern Recognition with Wearable Sensors". In: *Wearable Devices [Working Title]*. IntechOpen, June 2019. DOI: 10.5772/intechopen.85460. URL: https://doi.org/10.5772/intechopen.85460.

[82] Giorgia Acerbi et al. “A Wearable System for Stress Detection Through Physiological Data Analysis”. In: *Lecture Notes in Electrical Engineering*. Springer International Publishing, 2017, pp. 31–50. DOI: 10.1007/978-3-319-54283-6_3. URL: https://doi.org/10.1007/978-3-319-54283-6_3.

[83] Kumar Akash et al. “A Classification Model for Sensing Human Trust in Machines Using EEG and GSR”. In: *ACM Transactions on Interactive Intelligent Systems* 8.4 (Nov. 2018), pp. 1–20. DOI: 10.1145/3132743. URL: https://doi.org/10.1145/3132743.

[84] Joshua Poore et al. “Emulating Sociality: A Comparison Study of Physiological Signals from Human and Virtual Social Interactions”. In: Jan. 2012.

[85] Christian J. A. M. Willemse, Alexander Toet, and Jan B. F. van Erp. “Affective and Behavioral Responses to Robot-Initiated Social Touch: Toward Understanding the Opportunities and Limitations of Physical Contact in Human–Robot Interaction”. In: *Frontiers in ICT* 4 (May 2017). DOI: 10.3389/fict.2017.00012. URL: https://doi.org/10.3389/fict.2017.00012.

[86] Lewis R. Goldberg. “An alternative "description of personality": The Big-Five factor structure.” In: *Journal of Personality and Social Psychology* 59.6 (1990), pp. 1216–1229. DOI: 10.1037/0022-3514.59.6.1216. URL: https://doi.org/10.1037/0022-3514.59.6.1216.

[87] S.B.G. Eysenck, H.J. Eysenck, and Paul Barrett. “A revised version of the psychoticism scale”. In: *Personality and Individual Differences* 6.1 (Jan. 1985), pp. 21–29. DOI: 10.1016/0191-8869(85)90026-1. URL: https://doi.org/10.1016/0191-8869(85)90026-1.

[88] Pauline Chevalier et al. “Joint Attention using Human-Robot Interaction: Impact of sensory preferences of children with autism”. In: *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Aug. 2016. DOI: 10.1109/roman.2016.7745218. URL: https://doi.org/10.1109/roman.2016.7745218.

Appendix A

The BIG5 Personality Test

This is the list of questions comprised in the short, revised version of the BIG5 Personality Test that we have used for our experiments. Each participant had to answer, on a 5-point Likert scale, with 1 corresponding to "Strongly disagree" and 5 corresponding to "Strongly agree", how they consider and evaluate themselves regarding the following scenarios and habits:

1. ... is talkative.
2. ... tends to find fault with others.
3. ... does a thorough job.
4. ... is depressed, blue.
5. ... is original, comes up with new ideas.
6. ... is reserved.
7. ... is helpful and unselfish with others.
8. ... can be somewhat careless.
9. ... is relaxed, handles stress well.
10. ... is curious about many different things.
11. ... is full of energy.
12. ... starts quarrels with others.
13. ... is a reliable worker.
14. ... can be tense.
15. ... is ingenious, a deep thinker.
16. ... generates a lot of enthusiasm.
17. ... has a forgiving nature.
18. ... tends to be disorganised.
19. ... worries a lot.
20. ... has an active imagination.
21. ... tends to be quiet.
22. ... is generally trusting.
23. ... tends to be lazy.

24. ... is emotionally stable, not easily upset.
25. ... is inventive.
26. ... has an assertive personality.
27. ... can be cold and aloof.
28. ... perseveres until the task is finished.
29. ... can be moody.
30. ... values artistic, aesthetic experiences.
31. ... is sometimes shy, inhibited.
32. ... is considerate and kind to almost everyone.
33. ... does things efficiently.
34. ... remains calm in tense situations.
35. ... prefers work that is routine.
36. ... is outgoing, sociable.
37. ... is sometimes rude to others.
38. ... makes plans and follows through with them.
39. ... gets nervous easily.
40. ... likes to reflect, play with ideas.
41. ... has few artistic interests.
42. ... likes to cooperate with others.
43. ... is easily distracted.
44. ... is sophisticated in art, music, or literature.
45. ... is politically liberal.

Appendix B

The short, revised version of the Eysenck Personality Questionnaire

This is the list of questions comprised in the short, revised version of the Eysenck Personality Questionnaire that we have used for our experiments. The participants had to answer the questions by "Yes" or "No".

1. Does your mood often go up and down?
2. Do you take much notice of what people think?
3. Are you a talkative person?
4. If you say you will do something, do you always keep your promise no matter how inconvenient it might be?
5. Do you ever feel 'just miserable' for no reason?
6. Would being in debt worry you?
7. Are you rather lively?
8. Were you ever greedy by helping yourself to more than your share of anything?
9. Are you an irritable person?
10. Would you take drugs which may have strange or dangerous effects?
11. Do you enjoy meeting new people?
12. Have you ever blamed someone for doing something you knew was really your fault?
13. Are your feelings easily hurt?
14. Do you prefer to go your own way rather than act by the rules?
15. Can you usually let yourself go and enjoy yourself at a lively party?
16. Are all your habits good and desirable ones?
17. Do you often feel 'fed-up' ?
18. Do good manners and cleanliness matter much to you?
19. Do you usually take the initiative in making new friends?
20. Have you ever taken anything (even a pin or button) that belonged to someone else?
21. Would you call yourself a nervous person?
22. Do you think marriage is old-fashioned and should be done away with?

23. Can you easily get some life into a rather dull party?
24. Have you ever broken or lost something belonging to someone else?
25. Are you a worrier?
26. Do you enjoy co-operating with others?
27. Do you tend to keep in the background on social occasions?
28. Does it worry you if you know there are mistakes in your work?
29. Have you ever said anything bad or nasty about anyone?
30. Would you call yourself tense or 'highly-strung' ?
31. Do you think people spend too much time safeguarding their future with savings and insurances?
32. Do you like mixing with people?
33. As a child were you ever cheeky to your parents?
34. Do you worry too long after an embarrassing experience?
35. Do you try not to be rude to people?
36. Do you like plenty of bustle and excitement around you?
37. Have you ever cheated at a game?
38. Do you suffer from 'nerves' ?
39. Would you like other people to be afraid of you?
40. Have you ever taken advantage of someone?
41. Are you mostly quiet when you are with other people?
42. Do you often feel lonely?
43. Is it better to follow society's rules than go your own way?
44. Do other people think of you as being very lively?
45. Do you always practice what you preach?
46. Are you often troubled about feelings of guilt?
47. Do you sometimes put off until tomorrow what you ought to do today?
48. Can you get a party going?

Appendix C

Question tree used for the card guessing game

The following tables contain the question trees that were used in the card guessing game experimental setup to guess the card chosen by the participant based on their answers. Each question had two versions, with different phrasings, in order to render the discourse less monotonous, but the structure of the two question tables is identical. Advancing from one question to another is done depending on the answer given by the player to the previous question (**Y**es or **N**o), and for each unique series of 5 answers given to the 5 questions, there is exactly one possible image card corresponding to it.

TABLE C.1: Card guessing question tree: phrasing A

Question 1	Question 2	Question 3	Question 4	Question 5	Card image
Does your card show an object?	**Y:** Does your card show a vehicle?	**Y:** Does your card show a flying object?	**Y:** Does your card show a real flying object?	**Y:** Does the object on your card have wings?	**Y:** plane
					N: helicopter
				N: Does the object on your card have wings?	**Y:** X-Wing Starfighter
					N: The Millenium Falcon
			N: Does your card show a real vehicle?	**Y:** Does the object on your card have an engine?	**Y:** car
					N: bicycle
				N: Does the object on your card have wheels?	**Y:** Batmobile
					N: AT-AT transporter
		N: Does your card show an electronic device?	**Y:** Does your card show a modern electronic device?	**Y:** Does your card show a portable object?	**Y:** smartphone
					N: TV
				N: Does your card show a portable object?	**Y:** old radio
					N: old computer
			N: Does your card show a tool?	**Y:** Does your card show a heavy object?	**Y:** hammer
					N: screwdriver
				N: Does your card show a fragile object?	**Y:** vase
					N: bronze statue
	N: Does your card show a person?	**Y:** Does your card show a person who is now dead?	**Y:** Does your card show a man?	**Y:** Does your card show a scientist?	**Y:** Albert Einstein
					N: Elvis Presley
				N: Does your card show an actress?	**Y:** Marylin Monroe
					N: Amy Winehouse
			N: Does your card show a man?	**Y:** Does your card show a politician?	**Y:** Donald Trump
					N: Zinedine Zidane
				N: Does your card show an actress?	**Y:** Scarlett Johansson
					N: Serena Williams
		N: Does your card show a carnivore?	**Y:** Does your card show a being that lives in water?	**Y:** Does your card show a fish?	**Y:** shark
					N: crocodile
				N: Does your card show an animal with stripes?	**Y:** tiger
					N: wolf
			N: Does your card show a domestic animal?	**Y:** Does your card show an animal with horns?	**Y:** cow
					N: horse
				N: Does your card show an animal with stripes?	**Y:** zebra
					N: deer

TABLE C.2: Card guessing question tree: phrasing B

Question 1	Question 2	Question 3	Question 4	Question 5	Card image
Have you chosen an object?	**Y:** Have you chosen a vehicle?	**Y:** Have you chosen a flying object?	**Y:** Have you chosen a real flying object?	**Y:** Have you chosen an object with wings?	**Y:** plane
					N: helicopter
				N: Have you chosen an object with wings?	**Y:** X-Wing Starfighter
					N: The Millenium Falcon
			N: Have you chosen a real vehicle?	**Y:** Have you chosen an object that has an engine?	**Y:** car
					N: bicycle
				N: Have you chosen an object that has wheels?	**Y:** Batmobile
					N: AT-AT transporter
		N: Have you chosen an electronic device?	**Y:** Have you chosen a modern electronic device?	**Y:** Have you chosen a portable object?	**Y:** smartphone
					N: TV
				N: Have you chosen a portable object?	**Y:** old radio
					N: old computer
			N: Have you chosen a tool?	**Y:** Have you chosen a heavy object?	**Y:** hammer
					N: screwdriver
				N: Have you chosen a fragile object?	**Y:** vase
					N: bronze statue
	N: Have you chosen a person?	**Y:** Have you chosen a person who is now dead?	**Y:** Have you chosen a man?	**Y:** Have you chosen a scientist?	**Y:** Albert Einstein
					N: Elvis Presley
				N: Have you chosen an actress?	**Y:** Marylin Monroe
					N: Amy Winehouse
			N: Have you chosen a man?	**Y:** Have you chosen a politician?	**Y:** Donald Trump
					N: Zinedine Zidane
				N: Have you chosen an actress?	**Y:** Scarlett Johansson
					N: Serena Williams
		N: Have you chosen a carnivore?	**Y:** Have you chosen a being that lives in water?	**Y:** Have you chosen a fish?	**Y:** shark
					N: crocodile
				N: Have you chosen an animal with stripes?	**Y:** tiger
					N: wolf
			N: Have you chosen a domestic animal?	**Y:** Have you chosen an animal with horns?	**Y:** cow
					N: horse
				N: Have you chosen an animal with stripes?	**Y:** zebra
					N: deer

Titre : Détection du mensonge dans le cadre des interactions homme-robot à l'aide de capteurs et dispositifs non invasifs et mini invasifs

Mots clés : Robotique Sociale, Détéction du mensonge, Interaction Homme-Robot

Résumé : La Robotique Sociale met l'accent sur l'amélioration de l'aptitude des robots d'interagir avec les humains, y compris la capacité des premiers de comprendre les derniers. Les Robots Sociaux Assistants (RSA) visent à améliorer la qualité de vie de leurs utilisateurs à travers des interactions sociales. Les utilisateurs vulnérables, comme ceux qu'ont besoin de réhabilitation, thérapie ou assistance permanente, bénéficient le plus de l'aide des RSAs.
Une des responsabilités des RSAs est de s'assurer que leurs utilisateurs respectent les reccomendations therapeutiques et medicales, mais les utilisateurs humains ne sont pas toujours coopérants. Comme certaines études l'ont montré, les humains tentent parfois de tromper leurs robots assistants afin d'éviter de respecter leurs recommandations, la conséquence étant la dégradation de leur état de santé.
Cette thèse explore les manifestations physiologiques et comportementaux associés au mensonge dans le contexte de l'IHR, à partir de la recherche faite dans le cadre des interactions interhumaines. Compte tenu du fait que nous considérons qu'il est très important de ne pas détériorer la qualité de l'interaction de façon significative, notre travail se concentre sur l'évaluation de ces manifestations uniquement à l'aide des moyens et dispositifs non invasifs et mini invasifs, comme les caméras RGB, RGB-D et thermales, aussi bien que les capteurs portables.
À ce but, nous avons conçu des scénarios d'interaction in-the-wild pendant lesquelles les participants sont incités à mentir. Pendant ces expériences, nous surveillons et mesurons leur fréquence cardiaque, fréquence respiratoire, température de la peau, conductivité de la peau, l'ouverture de leurs yeux, la position et l'orientation de leur tête aussi bien que le temps de réponse aux questions. Nous avons trouvé des corrélations entre et la véracité des réponses des participants et les variations des paramètres susmentionnés. En plus, nous avons aussi étudié l'impact de la nature de l'interlocuteur (humain ou robot) sur les manifestations des participants.
Nous considérons que cette thèse et nos résultats représentent un grand pas en avant vers le développement de robots capables d'établir si leurs interlocuteurs sont honnêtes ou pas, et ainsi d'améliorer la qualité des IHRs et la capacité des RSAs d'exercer leurs fonctions et d'améliorer la qualité de vie de leurs utilisateurs.

Title : Lie detection in human-robot interactions using noninvasive and minimally-invasive devices and sensors

Keywords : Social Robotics, Lie Detection, Human-Robot Interaction

Abstract : Social Robotics focuses on improving the ability of robots to interact with humans, including the formers' capacity to understand the latter. Socially Assistive Robots (SAR) aim to improve the quality of life of their users by means of social interactions. Vulnerable users, like people requiring rehabilitation, therapy or permanent assistance, benefit the most from the aid of SARs.
One of the responsibilities of SARs is to make sure their users respect their therapeutic and medical recommendations, yet human users are not always cooperative. As studies have shown, humans sometimes deceive their robot caretakers in order to avoid following their recommendations, thus resulting in a deterioration of their health.
This thesis explores the physiological and behavioural manifestations associated to deception in HRI, based on previous research done in inter-human interactions. As we consider that it is highly important to not impair the quality of the interaction in any way, our work focuses on the evaluation of these manifestations by means of noninvasive and minimally-invasive devices, such as RGB, RGB-D and thermal cameras as well as wearable sensors.
To this end, we have designed several in-the-wild interaction scenarios during which participants are enticed to lie. During these experiments, we monitored their heart rate, respiratory rate, skin temperature, skin conductance, eye openness, head position and orientation, and their response time to questions. We found correlations between the veracity of the participants' answers and the evolution of these parameters. Moreover, we have studied the impact of the nature of the interlocutor (human or robot) on the participants' manifestations.
We believe this thesis and our results represent a major step forward towards the development of robots that are able to establish the honesty and trustworthiness of their interlocutors, thus improving the quality of HRI and the ability of SARs to perform their duties and to improve the quality of life of their users.

Institut Polytechnique de Paris
91120 Palaiseau, France