



HAL
open science

Exploring the relationship between morningness-eveningness, cognitive performance and the internal physiological state in different human-robot interaction scenarios

Roxana Agrigoroaie

► **To cite this version:**

Roxana Agrigoroaie. Exploring the relationship between morningness-eveningness, cognitive performance and the internal physiological state in different human-robot interaction scenarios. Robotics [cs.RO]. Université Paris Saclay (COMUE), 2019. English. NNT : 2019SACL005 . tel-02264141

HAL Id: tel-02264141

<https://pastel.hal.science/tel-02264141v1>

Submitted on 6 Aug 2019

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

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Exploring the relationship between morningness-eveningness, cognitive performance and the internal physiological state in different human-robot interaction scenarios

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'ENSTA-Paris

Ecole doctorale n°573 INTERFACES: approches interdisciplinaires, fondements,
applications et innovation
Spécialité de doctorat : Informatique

Thèse présentée et soutenue à Palaiseau, le 1er Juillet 2019, par

ROXANA AGRIGOROAIE

Composition du Jury :

Mme Silvia Rossi Professeur, Università degli Studi di Napoli Federico II	Rapporteure
M. Miguel Salichs Professeur, Universidad Carlos III	Rapporteur
M. David Filliat Professeur, ENSTA-ParisTech (U2IS)	Président du jury
Mme Ana Paiva Professeur, Instituto Superior Técnico, University of Lisbon	Examineure
M. Ryad Chellali Professeur, Nanjing Forestry University (MEC)	Examineur
Mme Adriana TAPUS Professeur, ENSTA-ParisTech (U2IS)	Directeur de thèse

Declaration of Authorship

I, Roxana AGRIGOROAIE, declare that this thesis titled, “EXPLORING THE RELATIONSHIP BETWEEN MORNINGNESS-EVENINGNESS, COGNITIVE PERFORMANCE AND THE INTERNAL PHYSIOLOGICAL STATE IN DIFFERENT HUMAN-ROBOT INTERACTION SCENARIOS” and the work presented in it are my own. I confirm that:

- This work was done wholly and 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 of all, I would like to express my gratitude to my PhD advisor Adriana TAPUS for proposing to do my PhD thesis under her supervision. I thank her for her support, her guidance, for helping me become the researcher that I am today and for encouraging me to do my best during these last few years and to meet the conference deadlines. I am grateful that I was able to attend so many conferences and at the same time that I was able to travel so much.

I would like to thank François Ferland for his help, support and for showing me the “proper way” of doing things. I am grateful to the social robotics group (current members, former members and interns): Thi-Hai-Ha Dang, Amir Aly, Pauline Chevalier, François Ferland, Arturo Cruz-Maya, Pierre-Henri Orefice, David Iacob, Chuang Yu, Mihaela Şoroştinean, Ştefan Ciocirlan, Radu Mihalcea for their support, the nice collaborations, and for the discussions regarding social robotics and not only.

I also would like to thank the members of U2IS at ENSTA-Paris for taking part in my experiments, for the many activities that we did together and for the nice atmosphere that was in the lab these last three and a half years.

I deeply thank my family for their support, their encouragement and for being there for me for the duration of this thesis. I would like to thank my husband for his support, his patience and for the agreeing to move to France so that I could pursue my ambition of doing this PhD.

Last but not least, I thank everybody who helped me during the completion of this thesis.

Contents

Acknowledgements	v
1 Introduction	1
1.1 Introduction	1
1.2 Motivation	2
1.3 Plan of the thesis	2
2 Related work	5
2.1 Introduction	5
2.2 Physiological activity in HRI	7
2.3 ME-type and cognitive performance	7
2.4 Physiological measures for cognitive load	8
2.4.1 Blinking	8
2.4.2 Galvanic Skin Response (GSR)	8
2.4.3 Facial temperature variation	9
2.5 Conclusion	9
3 Experimental Platforms	11
3.1 Introduction	11
3.2 Robotic platform: TIAGo	12
3.3 Robotic platform: Pepper	14
3.4 Sensors	15
3.4.1 Thermal camera	15
3.4.2 RGB-D sensor	16
3.4.3 GSR sensor	16
3.5 Cognitive games	16
3.5.1 Stroop Game	17
3.5.2 Matrix Task	17
3.6 Conclusion	18
4 Methodology	19
4.1 Introduction	19
4.2 Thermal data extraction and analysis:	19
4.2.1 Face detection	20
4.2.2 Facial feature point prediction	21
4.2.3 Thermal ROIs	22
4.2.4 Thermal data extraction	23
4.2.5 Thermal data analysis	23
4.3 Blinking	24
4.4 GSR	27
4.5 Questionnaires	29
4.5.1 Eysenck Personality Questionnaire (EPQ)	29

4.5.2	The Reinforcement Sensitivity Theory Personality Questionnaire (RST-PQ)	29
4.5.3	Morningness-Eveningness Questionnaire (MEQ)	30
4.5.4	Adult/Adolescent Sensory Profile Questionnaire (AASP)	30
4.6	Conclusion and Contribution	30
5	RQ1: Relationship between Cognitive Performance and Physiological response	33
5.1	Introduction	33
5.2	Experimental design	34
5.2.1	Robotic Platform and Sensors	34
5.2.2	Questionnaires	34
5.2.3	Scenario	35
5.2.4	Conditions	35
5.2.5	Participants	36
5.2.6	Interaction time	36
5.3	Results	37
5.3.1	Task performance based on condition parameters	38
5.3.2	Task performance based on user profile	39
5.3.3	Physiological parameters variation based on condition parameters	42
5.3.4	Physiological parameters variation based on user profile	43
5.3.5	Results for morning individuals	44
5.3.6	Results for evening individuals	45
5.3.7	Correlation results	47
5.3.8	Other results	48
5.4	Classification results	48
5.5	Discussion	51
5.6	Conclusions	52
5.7	Contribution	52
6	RQ2: Relationship between ME-type and time of the day in relation to cognitive performance	53
6.1	Introduction	54
6.2	Experimental design	54
6.2.1	Experimental platform	54
6.2.2	Questionnaires	57
6.2.3	Scenario and experimental task	58
6.2.4	Participants	59
6.2.5	Interaction time	60
6.3	Results	60
6.3.1	Stroop game	61
	Task performance based on different factors	61
	Results for morning type individuals	66
	Results for evening type individuals	67
	Results for intermediate type individuals	68
	Other results	69
6.3.2	Integer Matrix	71
	Task performance based on different factors	71
	Results for morning type individuals	73
	Results for evening type individuals	74
	Results for intermediate type individuals	74

Other results	74
6.3.3 Decimal Matrix	75
Task performance based on different factors	75
Results for morning type individuals	77
Results for evening type individuals	77
Results for intermediate type individuals	78
Other results	78
6.4 Discussion and the message to take home	78
6.5 Conclusion	79
6.6 My contribution	79
7 RQ3: Influence of Empathy, Emotional Intelligence and Fight/Flight system in HRI	81
7.1 Introduction	81
7.2 Experimental design	82
7.2.1 Robotic platform and sensors	82
7.2.2 Questionnaires	82
7.2.3 Participants	83
7.2.4 Scenario	83
Robot behavior	84
7.2.5 Hypothesis	86
7.3 Data extraction and analysis	86
7.3.1 GSR	86
7.3.2 Facial Temperature variation	87
7.4 Results	87
7.4.1 Panas questionnaire	87
7.4.2 GSR	90
7.4.3 Facial Temperature variation	93
7.5 Discussion	94
7.6 Conclusion	95
7.7 Contribution	95
8 Assistive applications (I)	97
8.1 Introduction	97
8.2 Related work	98
8.3 The ENRICHME project	99
8.3.1 Robotic system	100
8.3.2 Graphical User Interface	101
8.4 Lessons learnt from a first interaction with the elderly	102
8.4.1 Scenario	103
8.4.2 Interaction, data recording and analysis	103
8.4.3 Thermal data	105
8.4.4 Results and comments	106
8.4.5 First lessons learnt	107
8.4.6 Discussion	108
8.4.7 Conclusion	108
8.5 Results of a 5-day interaction scenario designed for the elderly: a pilot study	108
8.5.1 Scenario	108
8.5.2 Participant description	110
8.5.3 Physiological data extraction and analysis	110

8.5.4	Hypotheses	110
8.5.5	Results	110
	PANAS	111
	Digit Cancellation	111
	Integer Matrix Task	114
	Stroop game	115
	Hypothesis H1.	116
	Hypothesis H2.	117
	Other results	118
8.5.6	Discussion	118
8.6	ENRICHME project testing phase	119
8.6.1	RQ1: Which is the most used GUI application?	120
8.6.2	RQ2: Which is the most played cognitive game?	120
8.6.3	Performance analysis	120
8.6.4	Discussion	123
8.7	Conclusion	124
8.8	Contribution	124
9	Assistive applications (II)	125
9.1	Introduction	125
9.2	Literature review	126
9.2.1	Insomnia and cognitive performance	126
9.3	Experimental design	127
9.3.1	Questionnaires	127
9.3.2	Participants	127
9.3.3	Scenario	128
9.3.4	Data recording and analysis	130
9.4	Results	132
9.4.1	Performance results	132
	CPT Task	132
	Integer Matrix Task	134
	Stroop Task	134
	Discussion	134
9.4.2	Performance results for individuals with insomnia	135
	CPT Task	135
	Integer Matrix task	140
	Stroop Task	142
9.4.3	Physiological response analysis	143
	CPT Task	144
	Integer Matrix Task	152
	Stroop Task	154
9.5	Discussion	156
9.6	Conclusions	158
9.7	Contribution	158
10	Conclusion	159
10.1	General considerations	159
10.2	Thesis summary	159
10.2.1	Chapter 2: Related work	159
10.2.2	Chapter 3: Experimental Platforms	159
10.2.3	Chapter 4: Methodology	160

10.2.4 Chapter 5: RQ1 Relationship between cognitive performance and physiological response	160
10.2.5 Chapter 6: RQ2 Relationship between ME-type and time of the day in relation to cognitive performance	161
10.2.6 Chapter 7: RQ3 Influence of empathy, emotional intelligence and fight/flight system in HRI	162
10.2.7 Chapter 8: Assistive applications (I)	162
10.2.8 Chapter 9: Assistive applications (II)	163
10.3 Perspectives	163
A List of Publications	165
B Morningness Eveningness Questionnaire	167
C Regulatory Focus Questionnaire - proverb form	171
D Eysenck Personality Questionnaire	173
E Reinforcement Sensitivity Theory Personality Questionnaire	175
Bibliography	179

List of Figures

3.1	Social Robots	13
3.2	TIAGo robot	14
3.3	Pepper robot	15
3.4	(A) Optris PI640 thermal camera; (B) An example of an image captured with the Optris PI640 thermal camera	15
3.5	ASUS Xtion PRO RGB-D camera	16
3.6	GROVE GSR sensor	16
3.7	Stroop task (Agrigoroaie et al., 2017b) (A) Fixation phase; (B) Stimulus phase; (C) Inter-trial phase	17
3.8	Integer Matrix Task: (a) Easy level; (b) Difficult level	18
3.9	Decimal Matrix Task: medium level	18
4.1	Camera calibration process (Sorostinean et al., 2015)	21
4.2	Label image for object detection: (left) input image; (right) Rectangle for the face	21
4.3	Thermal face detection, feature points and ROIs	22
4.4	Temperature variation in the forehead region with the result of linear regression (the model was fitted with $r^2 = 0.85$ and $p < 2.2e-16$)	24
4.5	Offline blink detection steps: (top) Original data; (bottom) Filtered data using a Savitzky Golay filter	25
4.6	Offline blink detection steps: (top) Average values for each window and the threshold; (bottom) Detected valleys	26
4.7	(A) Facial feature points; (B) Eyes feature points (Agrigoroaie et al., 2017a)	26
4.8	Filtered GSR signal with the detected peaks	28
4.9	Ideal GSR signal with the computed features (Setz et al., 2010)	28
5.1	Experimental setup	35
5.2	Participants distribution based on ME-type and the time of the day when the task was performed	38
5.3	Reaction Time during congruent trials based on the extraversion trait of EPQ	40
5.4	Interaction plot for ME-type and the interaction time	41
5.5	AccGSR based on condition	42
5.6	AccGSR for morning individuals based on condition	45
5.7	RTCong for evening individuals based on the personality trait of extraversion	46
6.1	Registration page	55
6.2	User dashboard	56
6.3	Experiments page	56
6.4	Study statistics page	57

6.5	Reaction time during incongruent trials based on the time of the day for Stroop Game	62
6.6	Interaction plot for MIE-type and the interaction time for Stroop Game	64
6.7	Interaction plot for MIE-type and the interaction time for Stroop Game	65
6.8	Reaction time for incongruent trials for ME-M type individuals based on the time of the day for Stroop Game	66
6.9	Reaction time for incongruent trials for MIE-M type individuals based on the time of the day for Stroop Game	67
6.10	Reaction time for incongruent trials for MIE-I type individuals based on the time of the day for Stroop Game	68
6.11	Stroop game played based on BIG5-C personality trait	70
6.12	Interaction plot for MIE-type and the interaction time for Integer Matrix Game	72
6.13	Interaction plot for ME-type and the interaction time for Decimal Matrix Game	76
6.14	Time needed to solve a matrix based on the time of the day for ME-E type individuals during Decimal Matrix Game	77
7.1	(a) Participant placing a Jenga piece; (b) Experimenter knocks over the tower	84
7.2	Robot posture when cheering up the participants and knocking down the tower	86
7.3	Typical data for event based analysis. The vertical lines represent the bump events.	87
7.4	Distribution of the four GSR event based parameters: latency time, amplitude, rise time, recovery time	88
7.5	Example of ROIs	88
7.6	Positive mood difference based on the condition type (throw/no throw)	89
7.7	Latency based on RST-FFFS	90
7.8	Amplitude based on condition type	91
7.9	Rise time based on experimenter	91
7.10	Recovery time based on condition	92
7.11	Forehead temperature variation based on RST-FFFS type	94
8.1	GUI main menu	100
8.2	The robot in one of the user's home	103
8.3	The developed interface with the potential activities	104
8.4	The physical exercise feedback	104
8.5	The regions of interest shown for the first participant. Temperatures range from 20°C in dark purple to 40°C in light yellow.	106
8.6	Mouth temperature variation	107
8.7	(a) Experimental setup; (b) Facial feature points and ROIs	109
8.8	Digit Cancellation total game time over time: (a) Easy level; (b) Difficult level	112
8.9	Digit Cancellation: (a) Easy level; (b) Difficult level	113
8.10	Digit cancellation average time to make a selection based on the day	113
8.11	Integer Matrix correct answers: (a) Easy level; (b) Difficult level	114
8.12	Matrix task correctly solved matrices in the difficult level based on interaction time	115
8.13	Stroop Game average reaction time based on interaction day	115
8.14	H1 results: (a) Boxplot for AccGSR; (b) Boxplot for AvgHR	116

8.15	Integer Matrix Game: (a) AccGSR; (b) Number of GSR peaks	117
8.16	Participants interacting with the ENRICHME robot; (A) PUMS; (B) AKTIOS; (C) LACE	120
8.17	Example of Digit Cancellation	121
8.18	Performance evolution over time of participant P6 for Digit Cancellation	124
9.1	Experimental setup	130
9.2	Number of commissions based on the age group of the participants . . .	133
9.3	Average time to give an answer (RTTotal) based on the age group for the Stroop Task	135
9.4	HitRT based on presence of insomnia	136
9.5	HitRT based on age group	137
9.6	HitRT based on ME-Type	138
9.7	Interaction between sleep time and sleep quality on hitRT (C=Bad sleep quality; B=Medium sleep quality	139
9.8	Interaction between ME-type and interaction time on omissions	140
9.9	Correctly solved matrices based on difficulty and neuroticism level . . .	141
9.10	Time to solve a matrix based on difficulty and neuroticism level	141
9.11	Time to solve a matrix based on sleep time	142
9.12	Interaction between ME-type and the interaction time on the average time to solve a matrix	142
9.13	Average reaction time based on the ME-type	143
9.14	Variation of the AccGSR parameter over the six blocks depending on the session	144
9.15	Variation of the GSR average value over the six blocks depending on the age group	145
9.16	Forehead average temperature over the six blocks depending on the age group	146
9.17	Blinks based on sleep time	147
9.18	Blinks over the six blocks based on sleep time	147
9.19	Average temperature in forehead region based on ME-type during Medium Integer Matrix Task	153
9.20	Average temperature in periorbital region based on ME-type during Difficult Integer Matrix Task	153
9.21	Average GSR value based on the extraversion level during the Stroop Task	154

List of Tables

5.1	Participants distribution based on questionnaires results	37
5.2	Significant results for task performance based on user profile	39
5.3	Significant results for physiological parameters variation based on user profile	43
5.4	Significant results for morning individuals	44
5.5	Significant results for evening individuals	46
5.6	Significant results for correlations	47
5.7	Classification accuracy for individual physiological parameters	49
5.8	Classification accuracies for combinations of physiological parameters	50
6.1	Distribution of the participants	59
6.2	Distribution of the played games at different times of the day	60
6.3	Distribution of the played Stroop games at different times of the day based on ME-type and MIE-type	61
6.4	Significant results for task performance based on different factors for Stroop Game	62
6.5	Results for ME-E individuals for Stroop Game	67
6.6	Results for MIE-E individuals for Stroop Game	68
6.7	Results from pairwise comparisons based on age group for Stroop Game	69
6.8	Significant results for task performance based on different factors for Stroop Game	70
6.9	Distribution of the played Integer Matrix games at different times of the day based on ME-type and MIE-type	71
6.10	Results for ME-M individuals during Integer Matrix Game	73
6.11	Results for MIE-M individuals for Integer Matrix Game	73
6.12	Results for MIE-E individuals for Integer Matrix Game	74
6.13	Results for MIE-I individuals for Integer Matrix Game	74
6.14	Distribution of the played Decimal Matrix games at different times of the day based on ME-type and MIE-type	75
6.15	Results for ME-M individuals for Decimal Matrix Game	77
7.1	Participants distribution based on questionnaires	83
7.2	Participants distribution based on the conditions	85
7.3	Angles of arm joints of the robot when knocking down the tower	85
7.4	Results for the GSR parameters	92
7.5	Results for the AccGSR	93
8.1	Testing duration for each participant	119
8.2	Usage of each application by the participants	121
8.3	Number of times each game was played by the participants	122
9.1	Distribution of participants	128
9.2	Tasks performed by participants	129

9.3	Distribution of individuals with insomnia	136
9.4	Correlation results for the performance measures	148
9.5	Correlation results for the GSR parameters with the temperature and blink parameters	149
9.6	Correlation results for the forehead region parameters with the other physiological parameters	150
9.7	Correlation results for the nose region parameters with the other phys- iological parameters	151
9.8	Correlation results for the periorbital region parameters with the other physiological parameters	151
9.9	Correlation results for the perinasal region parameters with the other physiological parameters	152
9.10	Correlation results for the medium and difficult level of the Integer Matrix Task	155
9.11	Correlation results for the Stroop Task	156
B.1	Scores and ME-types	169

Chapter 1

Introduction

1.1 Introduction	1
1.2 Motivation	2
1.3 Plan of the thesis	2

1.1 Introduction

Robots are more and more present everywhere around us. They have seen a great increase in popularity in the last few decades. They have gone from the machines present only in factories, in controlled settings, isolated from the outside world, to being available in multiple shapes, sizes, and with multiple and distinct capabilities. Some are designed to be toys for children, others, are designed to be used as guides (in museums, research facilities, airports), while others are designed to be used as helpers in hospital settings.

What all these applications have in common, is that robots have to interact with humans. They are no longer isolated. Even in factories, the trend is to have robots designed to collaborate with human workers. For example, there is a public-private partnership funded by the European Union with the main purpose of designing the factories of the future¹. Their main goal is to develop a more sustainable and competitive industry. One of the key elements investigated is the role robotics will play in these factories of the future. According to *Robotics and AI improve factories of the future*, the role played by robotics is divided into six main directions: to potentially reduce operational costs, to increase productivity by being utilized 24/7, to minimize errors and delays, collaborative robots (COBOTS) can eliminate mundane tasks and can lead to an increase in employee productivity, to help improve the value chain, and last but not least, to lead to an enhanced customer experience. These robots are intended to interact directly and more importantly, safely with humans.

Going from factories, to other every-day situations, the authors of (Bartneck et al., 2004) provide the following definition for social robots: **a social robot is an autonomous or semi-autonomous robot that interacts and communicates with humans by following the behavioral norms expected by the people with whom the robot is intended to interact**. While, they are already becoming more and more popular, according to (Bartneck et al., 2004), their development will continue to escalate. Some of the questions that might arise from this definition are: How can robots interact and communicate with humans? Which are the behavioral norms that they should follow? Which are the important elements that should be considered? How should they behave when there is only one individual? Should it change when there are two individuals? What about a group of people? Furthermore, we have to keep in mind the place where the interaction takes place. It can happen

¹<https://www.effra.eu/factories-future>

in the shopping mall, in a school, in the airport, in a hospital, a research laboratory. We should consider the interaction place as well when designing the behavior of a social robot.

1.2 Motivation

We believe that one of the main things that research in the field of social robotics should focus on is the design of a natural interaction between a social robot and an individual. This interaction should take into consideration the profile of the individual, the emotional state, the mood, and specific preferences among others.

People are all very different. We have different personalities, preferences, sensory profiles, moods, etc. A social robot should take all these aspects into consideration when planning its behavior. However, this is not that simple. This difference that exists between humans is something very complex. It is not a trivial thing to model the profile of an individual. How an individual reacts and their expectations might change based on other factors as well (which are not related to the interaction); the person might have received a bad news, it might have a headache, it might be thinking of how to solve a different, not related problem, etc.

Therefore, the main purpose of this thesis is to investigate how such a user profile can be determined. Multiple psychological questionnaires are considered. The context chosen is represented by doing a cognitive task and determining how the emotional state of an individual changes in different human-robot interaction scenarios. Thus, the purpose of this thesis is to *explore the relationship between morningness-eveningness (ME), cognitive performance and the internal physiological state in different human-robot interaction scenarios.*

1.3 Plan of the thesis

The outline of this thesis is as follows:

- In Chapter 2, we present a literature review of the research directions related to this thesis: physiological activity in HRI, relationship between ME-type and cognitive performance, and physiological measures for cognitive load
- In Chapter 3, we present the experimental platforms used throughout this thesis: the robotic platforms (TIAGo and Pepper), the sensors (thermal camera, RGB camera, GSR sensor), as well as the cognitive games developed for this thesis (Stroop Game, Matrix Task)
- In Chapter 4, we present how we use the physiological sensors to measure and analyse different physiological signals (facial temperature variation, blinking, and GSR). Moreover, we present the psychological questionnaires used to define the user profile (Eysenck Personality Questionnaire, The Reinforcement Sensitivity Theory Personality Questionnaire, the Morningness-Eveningness Questionnaire and the Adult/Adolescent Sensory Profile Questionnaire)
- In Chapter 5, we investigate the relationship between cognitive performance and physiological internal state in a laboratory setting with the TIAGo robot displaying two different types of behavior (encouraging behavior and stressing behavior)

-
- A follow-up study was conducted to better understand the relationship between morningness-eveningness and the time of the day when a cognitive task is performed. This follow-up online study is presented in Chapter 6
 - In Chapter 7, we investigate the influence of empathy, emotional intelligence, and the fight/flight system in HRI. More specifically, we present a study where Pepper, spoils the Jenga game played by the participants and investigate their physiological response
 - In Chapters 8 and 9, we present the results from studies carried out with two distinct populations: interaction with the elderly, part of the EU Horizon2020 ENRICHME research project, and individuals suffering from sleep disorder, which are outpatients of the Sleep Unit from the University Hospitals Pitié Salpêtrière - Charles Foix in Paris
 - Chapter 10 concludes this thesis

Chapter 2

Related work

2.1	Introduction	5
2.2	Physiological activity in HRI	7
2.3	ME-type and cognitive performance	7
2.4	Physiological measures for cognitive load	8
	2.4.1 Blinking	8
	2.4.2 Galvanic Skin Response (GSR)	8
	2.4.3 Facial temperature variation	9
2.5	Conclusion	9

2.1 Introduction

For establishing natural every-day life interactions it is important to understand the internal state of the individual we are interacting with. If an individual is feeling sad, we can adapt our behavior accordingly and try to make the other person feel better. While this is true for human-human interaction, can we expect the same while interacting with a robot? Is a robot capable of reliably understanding the emotional state of the individual it interacts with? And can it use that information in an efficient way?

It was shown that the systems (either computer or robotic) that can detect the affective state of an individual (e.g., frustration, stress, anger) and that are capable of adapting according to these states are more likely to be perceived as more trustworthy (Cassell et al., 2000), persuasive (Reeves et al., 1996; Cruz-Maya et al., 2018), and natural (Nakatsu, 1998).

The adaptation of the behavior of a social robot to the individual it interacts with has been studied for some time in the field of human-robot interaction (HRI). The adaptation can be based on personality traits (Tapus et al., 2007a; Woods et al., 2007), the affective state of the individual (Liu et al., 2008), proxemics (Walters et al., 2005), or different body signals (Mitsunaga et al., 2006). Our previous work (Agrigoroaie et al., 2017b) has shown that the sensory profile of an individual can also be used.

In this thesis, we define the user profile as a function of personality (Eysenck et al., 1975; Corr et al., 2016; Goldberg, 1990), sensory profile (Brown et al., 2002), morningness-eveningness type (ME-type) (Horne et al., 1976), and other user preferences.

Many research papers from the social psychology literature show the existence of multiple theories of personality. In the field of HRI, the most used model is the Big5 model of personality (Goldberg, 1990). There are other models that are based on biological factors (Eysenck, 1991), or are based on neuroscience theory (Gray et al., 2000). The BIG5 questionnaire is based on the factorial statistical analysis, and

it measures five personality traits: Extroversion, Neuroticism, Agreeableness, Openness, and Conscientiousness. The Eysenck personality model (Eysenck et al., 1975) considers that personality traits have an underlying physiological basis. The four personality traits considered in the model are: Extroversion, Neuroticism, Psychoticism and Lie/Social desirability. Based on the Eysenck personality model, Gray proposed a new personality model, the Reinforcement Sensitivity Theory (Gray et al., 2000). This new model is based on the interaction of three neurobehavioral systems: the behavior activation system, the fight/flight/freeze system, and the behavior inhibition system. Further details are provided in Chapter 4.

As it was shown by Bandura in (Bandura et al., 1991) and (Bandura, 1999), the personal determinants and the physiological internal state of how humans are functioning, have an impact and influence on social behavior. Moreover, research has shown that the emotional state of an individual has an effect on different physiological parameters (Picard et al., 2001; Ioannou et al., 2014b).

One area in which little research has been done in HRI is the adaptation of the behavior of the robot based on the sensory profile of an individual. The Sensory Profile has been developed by Winnie Dunn (Dunn, 1994; Dunn, 1997) to measure the responses of children to everyday sensory experiences. Initially the model was developed only for children, but later research showed that it can be applied to adults as well (Brown et al., 2001). The model has been conceptualized on the basis of the relationship between a neurological threshold continuum (low and high) and a behavioral response continuum (accordance and counteract). Based on the intersection between the two continuum a four-quadrant model was developed (Dunn, 1997).

A low neurological threshold indicates that a low-intensity stimulus is required for an individual to react, while a high neurological threshold requires high-intensity stimulus or a longer time to react to the same stimulus. Accordance indicates that the behavior corresponds with the neurological threshold, while counteracting indicates that the person is responding contrary to the neurological threshold. The four quadrants are: Low Registration (high neurological threshold and accordance), Sensation Seeking (high neurological threshold and counteract), Sensory Sensitivity (low neurological threshold and accordance), and Sensation Avoiding (low neurological threshold and counteract). Each of these four quadrants is characterized by six sensory processing categories (i.e., taste/smell, movement, visual, touch, activity level, auditory).

One of the guidelines defined by Yanco in (Yanco et al., 2004) for designing interfaces for HRI is that the interface should lower the cognitive load on the user. By cognitive load it is understood the load that a particular task imposes on the performer (Paas et al., 1993). But, how can the cognitive load of an individual be measured? Can a robot determine the level of cognitive load?

In (Kramer, 1991), Kramer made a literature review of how cognitive load can be measured. According to his paper, human physiology reflects the changes in cognitive functioning. Therefore, the usage of different physiological parameters represents a viable solution. In (Kramer, 1991), Kramer reviews the following physiological parameters: the electroencephalographic activity (EEG), pupil diameter, endogenous eye blinks, event-related brain potentials, electrodermal activity (EDA), and cardiac activity. More recently, the authors of (Charles et al., 2019) have performed a literature review of the measures used for mental workload. They investigate previous works related to electrocardiac activity, blood pressure, ocular measures, respiration, skin based measures, and brain measures.

2.2 Physiological activity in HRI

In HRI, one study (Mower et al., 2007) investigated if physiological parameters can be used in order to accurately estimate the engagement level of an individual while playing a wire puzzle game moderated by a simulated or an embodied robot. By using GSR parameters and the skin temperature, the authors reached an accuracy of 84.73% in estimating the engagement level of the participants in the study. The authors concluded that such an information could enable the robot to adapt its behavior in order to re-engage the individual it interacts with.

In (Liu et al., 2008), the authors have used the cardiac activity, heart sound, bioimpedance, electrodermal activity, electromyographic activity and temperature of six children with Autism Spectrum Disorder for affective modeling. They were able to detect three emotional states: liking, anxiety and engagement. The average accuracy is 85% for liking, 79.5% for anxiety and 84.3% for engagement.

In a card game scenario developed for a humanoid robot, the authors of (Sorostinean et al., 2015), have used a thermal camera to measure the stress level of the individuals interacting with the robot. They have also found that the interaction distance between the robot and the individuals has an effect on the facial temperature. While being located in the personal space of the individuals (i.e., 0.6 - 0.7 m (Hall, 1966)), the temperature variation in the nose and perinasal regions was significantly higher than while being positioned in the social space (i.e., 1.2 - 1.3 m (Hall, 1966)). Furthermore, they have also found a significant interaction between the interaction distance and the type of gaze of the robot (i.e., direct, or averted).

By using the EEG signal of the participants in their study, the authors of (Ehrlich et al., 2014) have extracted two aspects of social engagement (i.e., intention to initiate eye contact, and the distinction between being the initiator or the responder of a gaze contact). These two measures enabled a robot to decide when and how to engage in an interaction with a human.

2.3 ME-type and cognitive performance

The morningness-eveningness type (ME-type) refers to the differences existing in individuals' preference for the optimum time of the day to perform a given activity (Adan, 1991). Based on this distinction, an individual can be part of one of three groups: morning type, evening type, and neither or intermediate type. The most widely used assessment tool for measuring the ME-type is the Morningness Eveningness Questionnaire (Horne et al., 1976). For the development of the questionnaire, the authors have evaluated the oral temperature of the subjects and its variation throughout the day for the three groups (i.e., morning, intermediate and evening types).

Further research has investigated the relationship between body temperature and efficiency. The research shows inconclusive results. Some (Fröberg, 1977) have found a negative correlation between the temperature and the omission errors for both morning and evening types. Others (Horne et al., 1980), have found a positive correlation for evening type individuals, and a negative correlation for the morning type individuals. For the study in (Horne et al., 1980), the authors have simulated a production-line inspection task and they found that morning type individuals had a better performance during the morning hours, while evening type individuals performed better in the evening hours. They also found that in the afternoon the differences between the two groups were smaller than during the morning hours.

In (Song et al., 2000) the authors have investigated the relationship between ME-type, time-of-day, and the speed for processing information. They used two types

of tests: inspection time test and a multidimensional aptitude battery. The authors found that morning type individuals performed significantly worse in the morning session for spatial sub-tests of the multidimensional aptitude battery. On the other hand, for the same task, evening type individuals performed significantly better in the morning hours than in the evening hours.

For a visual search task, the authors of (Natale et al., 2003) have found that morning type individuals were faster in the morning, while evening type individuals were faster in the afternoon hours.

2.4 Physiological measures for cognitive load

The response to a stimulus or a cognitive task can be measured by using certain physiological parameters (e.g., EEG, heart rate, respiration rate, GSR(Lisetti et al., 2004), facial temperature variation (Ioannou et al., 2014b), blinking, facial expressions). The measurement and the understanding of how these parameters vary are a good indicator of the internal state of an individual. Therefore, a robot that can measure these parameters in a contactless or a non-invasive way can determine the emotional state of the individual it interacts with. This could lead to the adaptation of a robot's behavior in order to better interact with the individual. The variation of these parameters is also dependent on different personality traits, the sensory profile, or if the individual is a morning or an evening type.

2.4.1 Blinking

In the context of HRI and our research questions, spontaneous blinking is of high interest. The blinking rate (BR) of an individual varies based on the activity performed. In (Bentivoglio et al., 1997), the authors have found that the average resting BR is of 17 blinks/minute, while in a conversation scenario, the BR increases to 26 blinks/minute. Furthermore, the authors of (Tanaka et al., 1993) found that the BR relates to the task and its difficulty level. The more difficult a task, the higher the BR. Moreover, blinking rate can also be influenced by exposing an individual to different visual or auditory stimuli (Stern et al., 1984; Fukuda, 2001). Exposure to the visual stimuli leads to a decrease in the frequency of the BR. Additionally, gender differences were found in BR. The authors of (Campagne et al., 2005) have found evidence that females displayed an increased BR compared to males. Two parameters are mostly used for blinking: blinking rate and the number of blinks (Nourbakhsh et al., 2013).

2.4.2 Galvanic Skin Response (GSR)

The GSR has been used for many years in the research of psychophysiology (Cacioppo et al., 2007). It was shown that it can be successfully used to determine the arousal level (Nebylitsyn et al., 1972), the cognitive load (Nourbakhsh et al., 2012), the emotional state of an individual (Lisetti et al., 2004), to differentiate between stress and cognitive load (Setz et al., 2010).

The study led by Setz (Setz et al., 2010) achieved a classification accuracy of 82.8% in a two-class classification of stress - no stress (i.e., cognitive load). They used both relative (i.e., the mean value recorded during the baseline was extracted from the features) and nonrelative GSR features (e.g., mean, maximum, minimum EDA level, mean EDA peak height, mean EDA peak rate). They found better performance for the non-relative features, indicating that baseline features are not mandatory for a good classification.

Moreover, the authors in (Nourbakhsh et al., 2012) used GSR to measure cognitive load in arithmetic and reading tasks. In (Nourbakhsh et al., 2013), the authors have successfully differentiated between different levels of cognitive load by using blink and GSR features. The features extracted and used for the classification are: the accumulative GSR (summation of GSR values over task time), GSR power spectrum, number of blinks, and blink rate. By combining the GSR power spectrum and the number of blinks, the authors have obtained a classification accuracy of 75%, in a 2-class classification problem by using a Naïve Bayes classifier. In their later study, (Nourbakhsh et al., 2017), the same group, has performed two studies in which they used six GSR features (number of peaks, peak amplitude, rise duration, peak area, accumulative GSR, and power spectrum) in order to differentiate between four levels of cognitive load. The common best features in the classification of either two or four cognitive load levels are the rise duration and the accumulative GSR (with classification accuracy higher than 80% for the two-class classification, and higher than 45% for the four-class classification).

2.4.3 Facial temperature variation

Another type of physiological feature used to measure cognitive load is represented by the facial temperature.

In their paper (Or et al., 2007), Or and Duffy, have developed a non-intrusive method for measuring the mental workload in a driving scenario by using the facial temperature variation. They found a significant correlation between the variation of the temperature in the nose region and the subjective workload score. An increase in the rating of the mental workload was correlated with a drop in the nose region. No differences were found for the forehead region.

Another study that investigated if the cognitive load can be estimated from the facial temperature variation is the study of Abdelrahman et al. (Abdelrahman et al., 2017). They considered two regions of interest (ROIs) (i.e., the nose and the forehead) during two tasks (i.e., Stroop test and a reading task). During both tasks the task difficulty was correlated with an increase in the forehead temperature and a decrease in the nose temperature.

In (Ioannou et al., 2014b), the authors have associated the emotional state of an individual with the variation of the facial temperature in different regions of interest. The authors of (Ioannou et al., 2014a) have performed a literature review to determine which are the most important ROIs on the face when considering the temperature variation. They found that the nose, the forehead, the chin, the cheeks, the periorbital regions, and the maxillary area provide the most useful information with regards to the internal state of an individual.

2.5 Conclusion

In this Chapter, we have presented some of the work that exists in the literature that is related to our own work. Next, in Chapter 3, we are going to present the experimental platforms used in this thesis.

Chapter 3

Experimental Platforms

3.1	Introduction	11
3.2	Robotic platform: TIAGo	12
3.3	Robotic platform: Pepper	14
3.4	Sensors	15
3.4.1	Thermal camera	15
3.4.2	RGB-D sensor	16
3.4.3	GSR sensor	16
3.5	Cognitive games	16
3.5.1	Stroop Game	17
3.5.2	Matrix Task	17
3.6	Conclusion	18

3.1 Introduction

As seen in Chapter 1, in this thesis we explore the relationship between morningness-eveningness, cognitive performance and the internal physiological state of an individual in different human-robot interaction scenarios. We showed in Chapter 2 that the internal physiological state can be measured using different physiological parameters (e.g., blinking, galvanic skin response, facial temperature variation). As we wanted to use, as much as possible, only sensors that can be mounted on a robotic platform, we decided to use as physiological sensors an RGB-D camera, a thermal camera, and a GSR sensor. We chose to use the GSR sensor, even if it needs to be placed on the fingers of the end-user, as, the GSR is a good indicator of the current internal state of an individual, of cognitive load, and of stress as well. By using these sensors, we want to enable a social robot to better understand what an individual feels during an interaction in general, and during a cognitive task, in particular.

Social robots are being widely used for research, teaching, and entertainment, among others. There are multiple types of social robots that exists. Below, we provide a short description of some of the most well known social robots (see also Figure 3.1).

- PARO (Wada et al., 2003)(see Figure 3.1a), is a robot developed in Japan at the Intelligent Systems Research Institute. It is shaped as a baby seal and it has a soft artificial fur. It is designed especially for the elderly and it was developed to study the effects of Animal Assistive Therapy (AAT) with companion robots. It was successfully used to reduce patient stress and their caregivers, to stimulate the interaction between patients and caregivers, and it can help improve socialisation. The robot features sensors for sight, hearing, and touch. While it is able to autonomously move its body, it is not a mobile robot.

- iCUB (Metta et al., 2008) (Figure 3.1c) was developed as part of the RobotCup collaborative EU project in Italy, for research in embodied cognition. Its dimensions are similar to that of a 3.5 year old child. The robot is open-source and it has capabilities such as: crawling, solving complex 3D mazes, expressing emotions through facial expressions, grasping small objects such as balls, plastic bottles.
- NAO¹ (Figure 3.1b) is a humanoid robot developed by Softbank Robotics, in France. It has a height of only 54 cm and a weight of 4.5 kg. It is used worldwide for research but it is also available in multiple schools, colleges and universities to teach programming and to do research in HRI.
- Robotis OP2² (Figure 3.1d) also known as Darwin2 is an open source miniature-humanoid robot (height 45cm; weight: 2.9kg). It is developed and manufactured by Robotis, in South Korea, in collaboration with the University of Pennsylvania. The robot was developed with advanced computational power. It is mostly used by universities and research centers for educational and research purposes.
- Meka M1 (Figure 3.1e) humanoid robot is a platform suited for research in HRI and social robotics. The robot features a wheeled base, a torso, an expressive head, and compliant arms. The mouth of the robot features an LED strip, which can be used to express different facial expressions. It was successfully used for different HRI experiments in multiple research centers. Due to its high cost, it is only used for research.

In our laboratory, we have the following social robots: Nao, TIAGo, Pepper, Kompai 1, Zeno, and Meka. We chose to use TIAGo as it is also the robotic platform used for the ENRICHME EU H2020 research project (most of the work in this thesis was conducted in the context of this research project), and Pepper, as it is able to display multiple gestures than TIAGo.

This chapter is divided into three main parts. In the first part, we describe the two robotic platforms used throughout this thesis. The second part is dedicated to the presentation of the sensors used for this thesis (i.e., RGB-D camera, thermal camera, GSR sensor). And the last part, presents the three cognitive games that were designed to be used as the cognitive tasks throughout the thesis.

3.2 Robotic platform: TIAGo

TIAGo³ (Pages et al., 2016) (see Figure 3.2) is a modular robot developed by PAL Robotics⁴ in Barcelona, Spain. The version of the robot used for this thesis is an IRON TIAGo that was customized based on the requirements of the EU H2020 ENRICHME research project.

The robot features a mobile base, a lifting torso, a touchscreen mounted on the torso, and a head.

The differential-drive mobile base contains an on-board computer, a laser range-finder, three rear sonars, and an Inertial Measurement Unit, and it has two degrees of freedom. The maximum speed for the base is of 1m/s.

¹<https://www.softbankrobotics.com/emea/en/nao>

²http://emmanual.robotis.com/docs/en/platform/op2/getting_started

³<http://tiago.pal-robotics.com>

⁴<http://www.pal-robotics.com>

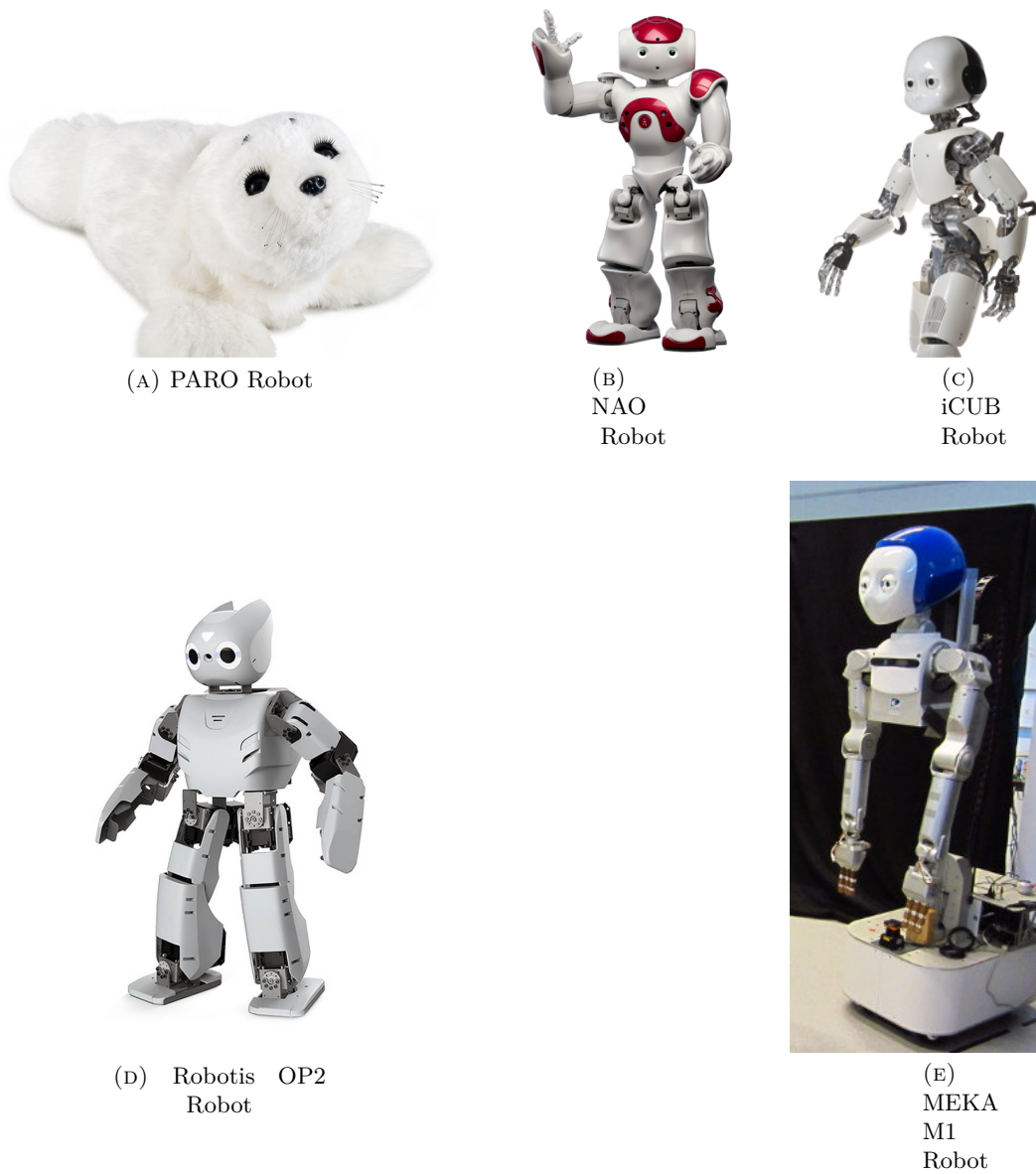


FIGURE 3.1: Social Robots

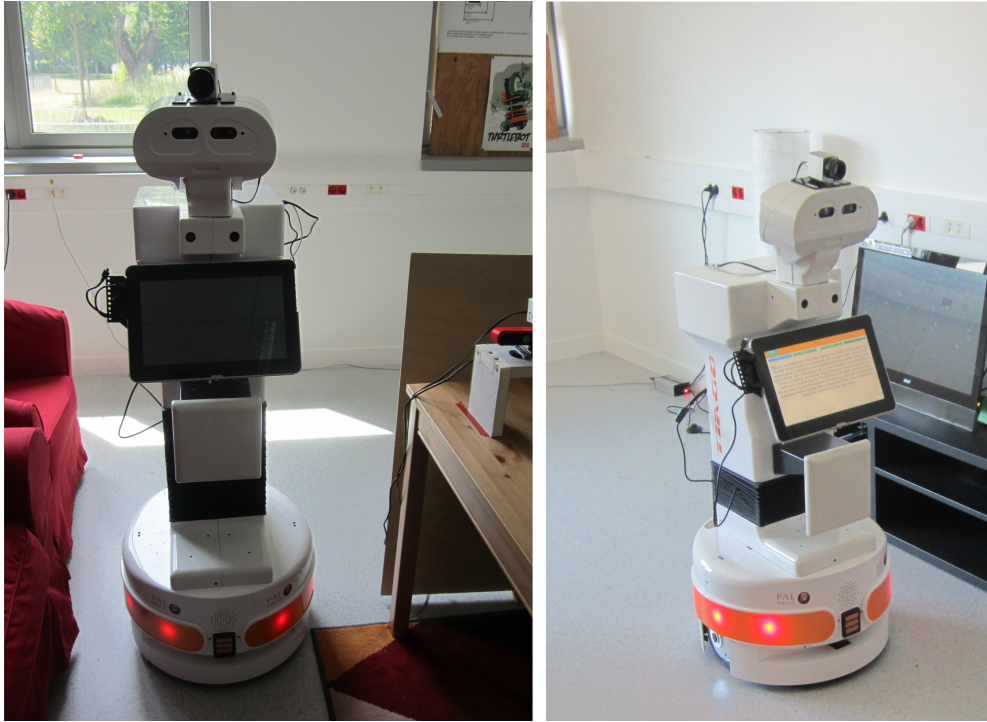


FIGURE 3.2: TIAGo robot

The torso is equipped with an internal lifter mechanism which allows changing the height of the robot (between 110 and 145 cm). On the torso of the robot a touchscreen is also mounted (GeChic OnLap 1303i⁵). On top of the robot (just behind the head) there is also a laptop tray which can support weights of up to 5kg.

The head is equipped with a pan-tilt mechanism (i.e., 2 degrees of freedom). Other sensors included in the head of the robot are an Orbbec Astra RGB-D⁶ camera, two stereo microphones and a speaker. On top of the head an USB-powered Optris PI460⁷ thermal camera was mounted as well.

The robot features a navigation system that can be used to perform the mapping as well as to enable the robot to autonomously navigate on a given map. The mapping system of the robot (i.e., GMapping⁸) uses the readings from the 2D laser scanner, which is located on the mobile base to create an Occupancy Grid Map (OGM)(Elfes, 1989). The robot is 100% ROS based.

The robot is integrated with the Acapela⁹ voice synthesis software for which a ROS interface was created.

3.3 Robotic platform: Pepper

Pepper¹⁰ (see Figure 3.3) was developed by SoftBank robotics and is considered to be the first social robot that is capable of recognizing faces and basic human emotions.

⁵<https://www.gechic.com/en-portable-touch-monitor-onlap1303i-view.html>

⁶<https://orbbec3d.com/product-astra-pro/>

⁷<https://www.optris.global/thermal-imager-optris-pi-640>

⁸<http://wiki.ros.org/gmapping>

⁹<https://www.acapela-group.com/>

¹⁰<https://www.softbankrobotics.com/emea/en/pepper>



FIGURE 3.3: Pepper robot



FIGURE 3.4: (A) Optris PI640 thermal camera; (B) An example of an image captured with the Optris PI640 thermal camera

The robot with a height of 120 cm, features 3 omni-directional wheels, a curved torso, a touchscreen mounted on the torso, and a head. The robot operating system is NAOqi¹¹. The maximum moving speed is of 1.4 m/s.

The head of the robot features 4 microphones, 2 RGB HD cameras (one in the mouth zone and one in the forehead). Furthermore, the robot also features 5 tactile sensors (three of them on the head, and one on each hand). For depth perception, a 3D camera is mounted in the eyes of Pepper. The base includes 2 sonars, 6 lasers, and one gyro sensor. Text-to-speech and speech recognition software for 15 languages are also included.

3.4 Sensors

3.4.1 Thermal camera

For this thesis the Optris¹² PI640 (see Figure 3.4a) USB-powered thermal camera was used. The camera is one of the smallest thermal cameras (45x56x90 mm) available on the market with a weight of only 320 grams.

With a frame-rate of 32Hz and a spectral range of 7.5 to 13 μm the camera has an optical resolution of 640x480 pixels. An example image taken with the camera is shown in Figure 3.4b.

¹¹http://doc.aldebaran.com/2-5/index_dev_guide.html

¹²<https://www.optris.global/thermal-imager-optris-pi-640>



FIGURE 3.5: ASUS Xtion PRO RGB-D camera

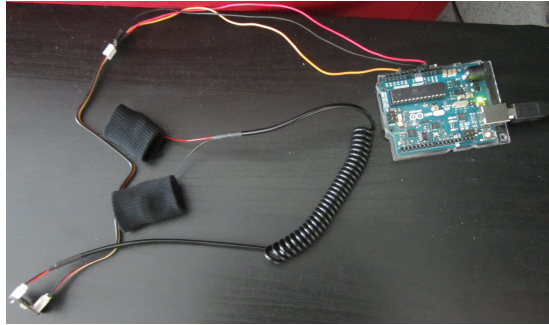


FIGURE 3.6: GROVE GSR sensor

3.4.2 RGB-D sensor

The RGB-D sensor used for this thesis is the ASUS Xtion PRO¹³ (see Figure 3.5). The USB-powered camera provides an optical resolution of 640x480 pixels at a frame rate of 30 fps. The camera provides RGB, Depth as well as pose estimate (22 joint positions) with the OpenNI body tracker.

3.4.3 GSR sensor

The galvanic skin response sensor used throughout this thesis is the GROVE GSR sensor¹⁴ (see Figure 3.6).

The analog output of the sensor can be converted to skin resistance by using the conversion formula provided by the manufacturer (see Equation 3.1). The output of the sensor ranges between 0 and 1023.

The sensor is usually placed on the fingers of the participant, on the ring and middle fingers of the non-dominant hand. In order to be used, the sensor can be easily connected to an Arduino board or a Raspberry Pi board. For this thesis, we have used an Arduino Uno board and have created a ROS module to publish the skin resistance data.

$$skin_R = \frac{(1024 + 2 * sensor_reading) * 1000}{512 - sensor_reading} \quad (3.1)$$

3.5 Cognitive games

For this thesis some cognitive games were developed to be used in the different experiments. The cognitive games developed are: the Stroop game, and the Matrix task (both integer and decimal version). Next, each game is briefly described.

¹³https://www.asus.com/3D-Sensor/Xtion_PRO

¹⁴http://http://wiki.seeedstudio.com/Grove-GSR_Sensor/

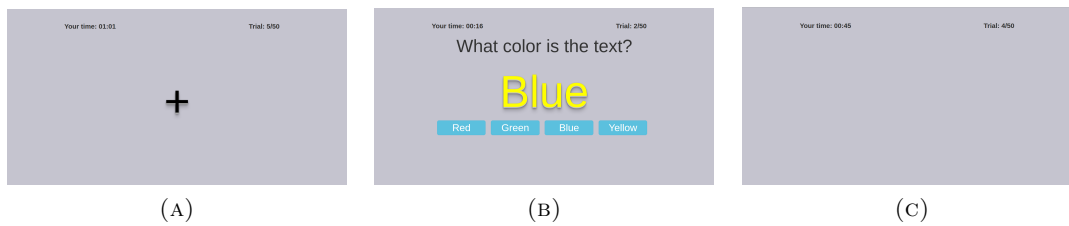


FIGURE 3.7: Stroop task (Agrigoroaie et al., 2017b) (A) Fixation phase; (B) Stimulus phase; (C) Inter-trial phase

3.5.1 Stroop Game

For this game, a variation of the original word-color Stroop test (Stroop, 1935) was developed. The game has been previously used as an experimental stressor (Renaud et al., 1997; Svetlak et al., 2010; Tulen et al., 1989) and no significant differences were found between males and females (MacLeod, 1991).

The user has to press on the button, which corresponds to the color of the word written on the screen (“What color is the text?”). The order of the buttons is randomized between trials. The game was developed with four possible colors: Red, Green, Blue, and Yellow. For each button press a feedback is given. Two versions were developed: one with no time limit and one with 3 second time limit. Each trial consists of three phases: the fixation phase, the trial phase, and the inter-trial phase (see also Figure 3.7). Before each stimulus was shown, a warning signal was given in the form of a fixation phase of 1 second. The purpose of this phase was to prepare the participant for the stimulus and to lead to a reduction of the reaction time (Kahneman, 1973).

The number of total trials of this game can be set as a parameter (we chose a trial number equal to 50). In this game there were two types of trials: congruent, and incongruent. In a congruent trial the color of the text is the same as the color depicted by the word written on the screen (e.g., the word **Blue** written in blue color). For the incongruent trial the color of the text was not the same as the color depicted by the word (e.g., the word **Blue** written in red color). The congruent and incongruent trials were presented in a random order.

The parameters of interest for this game are: the average reaction time for congruent trials, the average reaction time for incongruent trials, the average reaction time for all trials, correct answers, wrong answers, and number of trials with no answer.

3.5.2 Matrix Task

In this game the user is shown a matrix with digits. The goal is to find the two digits whose sum equals 10. Feedback is given after each selection. Once the matrix is correctly solved, a new one is presented. If the user makes a mistake, a new matrix is shown. The final goal is to solve as many matrices as possible in a total time of 5 minutes (the time limit can be set as a parameter). Multiple difficulty levels (see Figure 3.8) were developed: easy (Figure 3.8a) (matrix size of 2 x 2), medium (matrix size of 3x3), and difficult (Figure 3.8b) (matrix size of 4 x 4).

Two variations of this game were developed: Integer matrix task (with digits depicting integers) and Decimal matrix task (with digits depicting decimals with one decimal point, see Figure 3.9). The parameters of interest for this game are: the time needed to find the first digit, the time needed to find the second digit, the time

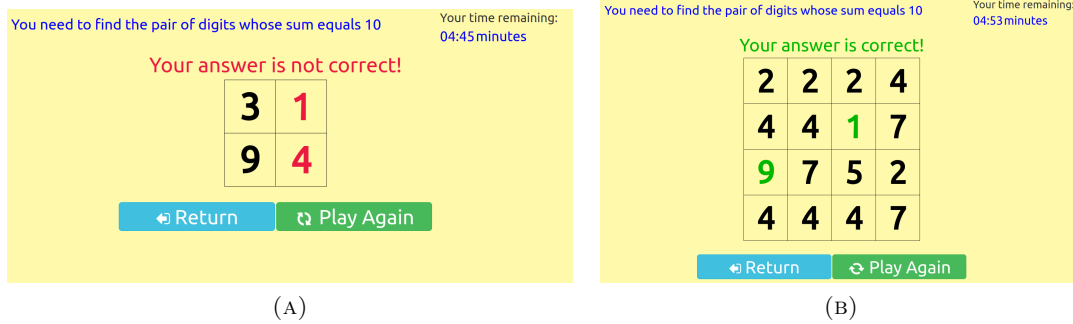


FIGURE 3.8: Integer Matrix Task: (a) Easy level; (b) Difficult level

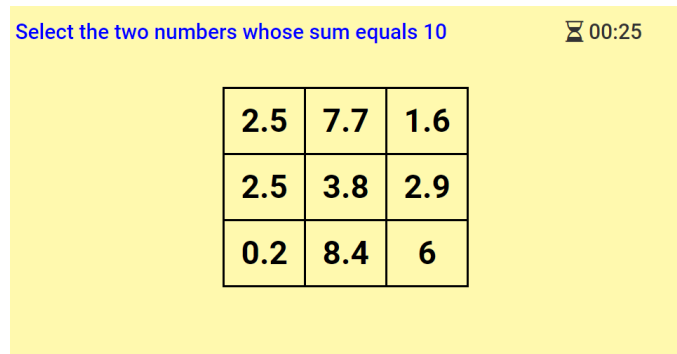


FIGURE 3.9: Decimal Matrix Task: medium level

needed to solve each matrix, the number of correct matrices, and the number of mistakes made.

3.6 Conclusion

In this Chapter we have presented the robotic platforms, as well as the physiological sensors used for this thesis. The two robotic platforms and the physiological sensors were used for carrying out the experiments presented in the following chapters. Furthermore, we have also presented the three cognitive games that were used as the cognitive tasks investigated in the following chapters.

Chapter 4

Methodology

4.1	Introduction	19
4.2	Thermal data extraction and analysis:	19
4.2.1	Face detection	20
4.2.2	Facial feature point prediction	21
4.2.3	Thermal ROIs	22
4.2.4	Thermal data extraction	23
4.2.5	Thermal data analysis	23
4.3	Blinking	24
4.4	GSR	27
4.5	Questionnaires	29
4.5.1	Eysenck Personality Questionnaire (EPQ)	29
4.5.2	The Reinforcement Sensitivity Theory Personality Questionnaire (RST-PQ)	29
4.5.3	Morningness-Eveningness Questionnaire (MEQ)	30
4.5.4	Adult/Adolescent Sensory Profile Questionnaire (AASP)	30
4.6	Conclusion and Contribution	30

4.1 Introduction

In the related work chapter (Chapter 2), we showed that for a natural interaction between a robot and an individual, the robot needs to be able to determine the internal state of that individual. For this, multiple physiological signals can be used: heart rate (HR), respiration rate, galvanic skin response (GSR), facial expressions, blinking, and facial temperature variation among others. For this thesis, we have decided to focus on three of these signals. By using the sensors presented in Chapter 3, the following physiological parameters are investigated: the facial temperature variation (using the thermal camera), blinking (using the RGB-D camera), and the galvanic skin response (using the Grove GSR sensor).

This Chapter is divided into two main parts. In the first part, we are going to present the methodology for the detection, extraction, and analysis of the physiological parameters. While in the second part, we are going to present the psychological questionnaires that enable us to determine the profile of an individual. More specifically, we are going to briefly describe the questionnaires, the traits that they measure and how each trait is computed.

4.2 Thermal data extraction and analysis:

As presented in Chapter 2, the temperature variation across different regions of interest (ROI) on the face provides a good insight into the current internal state of an

individual. For a robotic system to be able to use the thermal information, a method had to be proposed, for this thesis, for the extraction and analysis of this data. By using a thermal camera, the first step consists in the detection of faces in thermal images. Knowing the location of the face in the image, the next step consists in the prediction of some key facial feature points on the face (e.g., the corners of the eyes, the corners of the mouth). Based on these points multiple regions of interest can be defined on the face. In Chapter 2 (Related work) we have shown that the most important regions on the face are: the left, and right periorbital regions, the forehead, the left, and right cheek, and the nose, among others. Next, from each of these regions the data needs to be extracted and analysed before a robotic system can use it. Therefore, the five main steps are:

1. Face detection
2. Facial feature point prediction
3. Defining the ROIs
4. Temperature data extraction
5. Temperature data analysis

Next, each of the five steps is detailed.

4.2.1 Face detection

For the detection of faces in RGB images, there are multiple algorithms already implemented in different libraries, for example, OpenCV or Dlib (King, 2009). We first tried using these algorithms on our thermal images. However, they did not prove to be successful.

As we could not detect faces in RGB images, our next solution was to use a calibration process between an RGB camera and the thermal camera. For this, we used the camera calibration¹ ROS module. In order to perform this calibration (i.e., determining the geometrical relationship between the two cameras), we first had to fix the relative position between the two cameras. If the position was changed, the calibration had to be redone. Once the two cameras were positioned, we used a chessboard pattern with black and white squares. The pattern could easily be detected in RGB images, while for the thermal camera, the black squares had to be heated, and the white squares had to be cooled down. Using this process (also illustrated in Figure 4.1), we could more reliably detect faces in thermal images, than by using the face detection algorithms from the previously mentioned libraries. However, this process is not very practical, as the calibration process has to be redone very frequently.

Our next solution was to search for an already available face detector for thermal images, however we could not find any such detector. While using the Dlib library for the detection of faces in RGB images, we found that the library offers the possibility of training a custom object detector. Dlib² is an open source toolkit which contains machine learning algorithms and tools that are free to use to solve real-world problems. The library was originally written only in C++, but later it included implementation of some of the algorithms for Python as well. The face detection algorithm implemented in the library uses Histogram Oriented Gradients (HOG) (Dalal et al.,

¹http://wiki.ros.org/camera_calibration

²<http://dlib.net/>

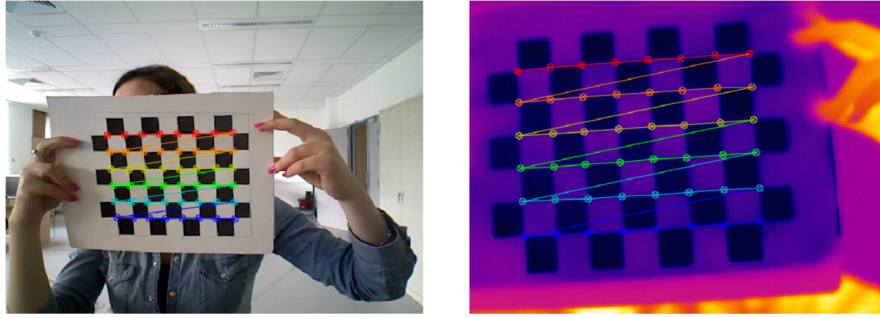


FIGURE 4.1: Camera calibration process (Sorostinean et al., 2015)

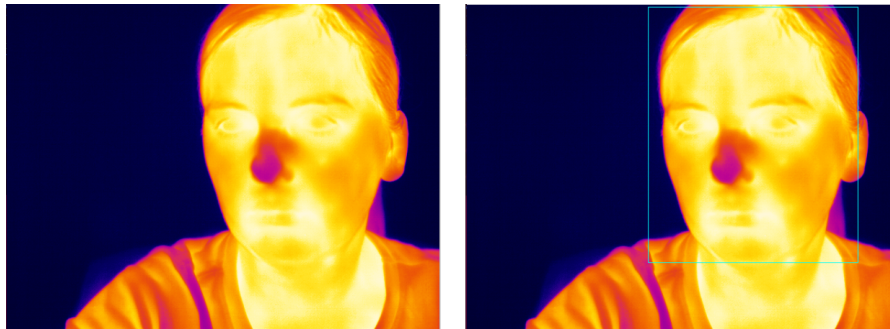


FIGURE 4.2: Label image for object detection: (left) input image; (right) Rectangle for the face

2005) features combined with a linear classifier. This type of object detector can be used to detect multiple type of objects, not just faces.

Based on the training of a new object detector method used in Dlib, we have developed a new Python module that enables the training of a face detector for thermal images. The current detector (available on Bitbucket³) was trained with more than 2000 images from the Natural Visible and Infrared Facial Expression Database (NVIE)(Wang et al., 2010) as well as images taken during the different experiments carried out in our laboratory. We used images with and without glasses as well as images taken from multiple distances from the camera. This enables us to detect thermal faces at different interaction distances. To train the detector, for each image a rectangle is drawn with the mouse around the face (see Figure 4.2). The location of the face is automatically recorded into an .xml file, in the format required by the training algorithm of Dlib (i.e., for each face the coordinates of a box are saved: height, width, and the coordinates of the top left corner of the box). When all images have been labeled, the individual .xml files are concatenated and the training can be performed. In its current format, the images are taken in batches of 10 (so that if a mistake is made, only the images in the current batch need to be re-labeled). A graphical user interface (GUI) is going to be designed for easier use of the software.

4.2.2 Facial feature point prediction

Once the face is detected, in order to define different ROIs on the face, some facial feature points have to be detected. The facial feature point predictor of the Dlib toolkit was trained on the iBUG 300-W dataset (Sagonas et al., 2016) and can localise 68 feature points (see Figure 4.7a) in RGB images. As with the face detection, the

³https://bitbucket.org/roxana_agr/

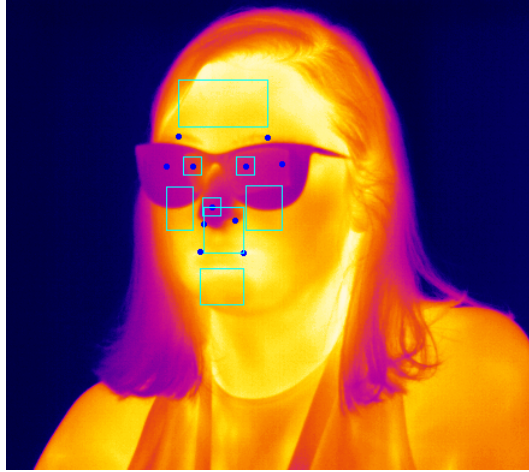


FIGURE 4.3: Thermal face detection, feature points and ROIs

facial feature point predictor did not work for thermal images. Thus, a facial feature point predictor was also trained to localize points of interest on a face. The shape predictor, which uses regression trees, is based on the algorithm presented in (Kazemi et al., 2014) and is part of the Dlib toolkit. We have decided that using 11 points of interest is sufficient in order to define multiple ROIs. The feature points are: the middle of the eyebrows, the inner and outer corners of the eyes, the corners and the tip of the nose, and the corners of the mouth. The facial feature point predictor (available on Bitbucket⁴) was trained on the same number of images as the detector. For the training, the images are split into chunks of 10 images (in case the user makes a mistake, only the 10 images need to be re-labeled). As with the face detector, the position of each of the 11 feature points is recorded into .xml files. These, are then used to train the predictor. An example of the detected facial feature points of interest are shown in Figure 4.3.

4.2.3 Thermal ROIs

For this thesis we have defined 8 ROIs for thermal images. As an interaction between the robot and the participant can take place at multiple distances from the thermal camera, the ROIs have to be dependent on the interaction distance. We chose to use the distance between the inner corners of the eyes (named *eyes_dist*) as the reference. Using the 11 feature points, the ROIs used in this thesis are (shown in Figure 4.3):

- **the forehead region:** width equal to the distance between the middle of the eyebrows; and the height equal to the distance between the eyes and the nose.
- **the left, and right periorbital regions:** both regions were defined as square regions around the inner corners of the eyes with the side equal to $1/3$ of *eyes_dist*.
- **the nose:** a square region around the tip of the nose, with the side equal to $1/3$ of *eyes_dist*
- **the perinasal region:** width equal to the distance between the corners of the mouth; and a height equal to the distance between one corner of the nose and

⁴https://bitbucket.org/roxana_agr/

the mouth at which a distance of $1/3$ of `eyes_dist` was added in order to include the nostrils too.

- **the chin region:** width equal to the distance between the corners of the mouth, and the height equal to `eyes_dist`
- **the left, and right cheek regions:** rectangular regions with the width equal to the length between the corners of the eyes, and the height equal to the distance between the midpoint of the corner of the eyes and nose, and the corner of the mouth and nose

In the work of (Ioannou et al., 2014a), the authors have found that the most important regions on the face that provide insight into how an individual is feeling are: the nose, cheeks, periorbital region, chin, maxillary area, and the forehead. However, there are no indications as to how to define these regions. Therefore, we have decided to define them as presented above.

4.2.4 Thermal data extraction

The parameter mostly considered for the facial temperature variation is the rate of change (slope) (Sorostinean et al., 2015). Other parameters that can be considered are: the temperature range (i.e., $max_{temperature} - min_{temperature}$) during the interaction and the average temperature during the interaction. The data extraction can take place both in real-time, as well as offline, from ROS bag files. The thermal face detection and facial feature predictor, as well as the facial regions of interest are currently integrated with ROS, meaning that the coordinates of each region are available as ROS messages. For the real-time extraction, the average temperature from each ROI is extracted and either published as a ROS message, or saved in an .csv file.

For the offline extraction, all the temperatures from the ROIs are extracted and saved in a .csv file. For compression reasons, for each frame, all unique temperatures are extracted, together with how many times each one appears in the region. Next, during the analysis, all temperatures which are not associated to the face have to be eliminated. For example, there might be regions which contain glasses (i.e., periorbital regions, and sometimes the forehead too). These temperatures are usually below 30° Celsius (Agrigoroaie et al., 2016), therefore, a threshold can be easily applied to eliminate all temperatures below 30°C . Next, the average temperature from each region can be extracted and used for the analysis.

4.2.5 Thermal data analysis

Once the ROIs were defined, the mean temperature could be extracted together with the timestamp at which it occurred. In order to remove short variations that could lead to false measurements we re-sampled the data with linear interpolation with a rate of 10 Hz, and we applied a moving average filter with a window size of 20 samples (5 seconds). Finally, to fit a linear model on our data, a least-square regression was applied. An example of filtered forehead temperature with the result of linear regression (the model was fitted with $r^2 = 0.85$ and $p < 2.2e-16$) can be seen in Figure 4.4. In this case, the temperature increases with 0.017°C/s .

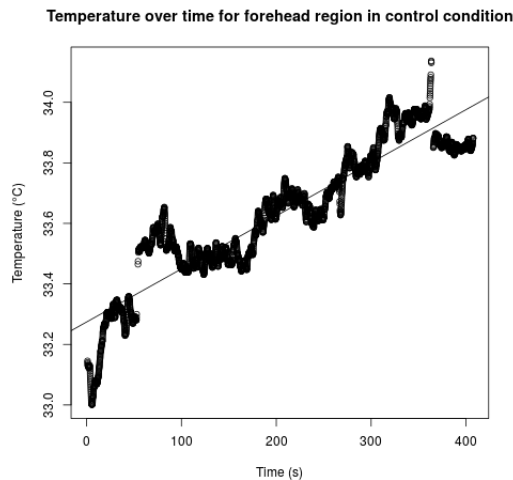


FIGURE 4.4: Temperature variation in the forehead region with the result of linear regression (the model was fitted with $r^2 = 0.85$ and $p < 2.2e-16$)

4.3 Blinking

We have shown in Chapter 2 that blinking is dependent on the task performed. As in this thesis we are interested in how the physiological internal state of an individual differs based on the task performed, blinking parameters should also be considered.

In order to detect the blinks, the RGB-D camera was used. We have developed a blink detection algorithm (Agrigoroaie et al., 2017a) based on the distance between the eyelids, which can be applied both online and offline. For both methods, we start by applying the face detection and facial feature point prediction algorithms from the Dlib toolkit. As seen in the previous section, these were trained on RGB images from the iBUG300-W dataset (Sagonas et al., 2016).

The facial feature point predictor gives us the location of 68 feature points of interest (see Fig. 4.7a). In order to determine if a person blinks or not, we look at the distance between the eyelids. As can be seen in Figure 4.7b, each eye is characterised by six points of interest: the corners of the eyes, and two points for each eyelid. Therefore, for each eye, we compute the Euclidean distance between the upper and lower eyelids for both points (e.g., in Figure 4.7b the Euclidean distance between the feature points 37 and 41).

When a person is very close to the camera the face region is very large, but as the distance between the camera and the person increases, the face region decreases. Therefore, the distance between the eyelids can be very small. To counteract this, we squared the sum of the two distances for each eye. Considering that there are multiple eye shapes and sizes, we need to be able to adapt the algorithm for each person. For this purposes, we consider a duration of 60 seconds (considered as a parameter that can be changed) in which we record the eyelid distances for both eyes. At the end of this period, we compute the mean eyelid distance for each eye. This mean value will enable us to determine if a blink occurs or not.

For the online analysis, we consider that an individual has its eyes closed if the current eyelid distance is smaller than 0.3 times the mean distance for that eye. This threshold was found empirically. If the person reopens his/her eyes in less than 300 ms we consider that a blink occurred, otherwise, we consider that the person had its

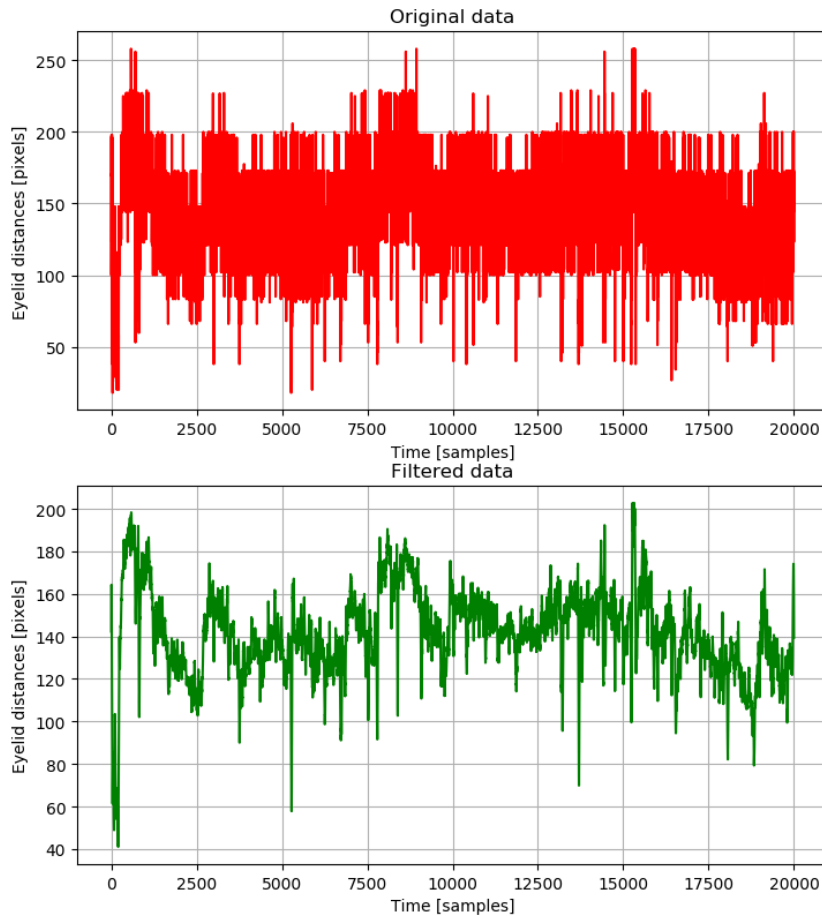


FIGURE 4.5: Offline blink detection steps: (top) Original data; (bottom) Filtered data using a Savitzky Golay filter

eyes closed. We consider that a blink happened only if a blink occurred for both eyes, otherwise, we consider that the person performed a wink.

For the offline analysis, we consider all eyelid distances, which are saved in .csv files (in Figure 4.5 top is shown the original signal). On the saved values, we first apply a Savitzky Golay (Savitzky et al., 1964) smoothing filter (see Figure 4.5 bottom for the filtered signal) so as to eliminate small variations in the signal. We applied the filter available in the Scipy⁵ Scientific Computing Python library. Next, for window sizes of 5 seconds (with a sampling frequency of 30Hz, a window size of 150 samples) we determine the average eyelid distance (see Figure 4.6 top). All valleys, which are lower than 0.95 times the average value, are recorded (see Figure 4.6 bottom). We apply this windowing, as the participants can move closer and further away from the RGB camera. If, at the same timestamp, valleys are detected in both the left and the right eyes, we consider that the participant blinked.

The blink detector works as a ROS module. For this thesis, the data was analyzed only offline. One limitation of our blink detector is that we cannot detect blinks if

⁵<http://www.scipy.org>

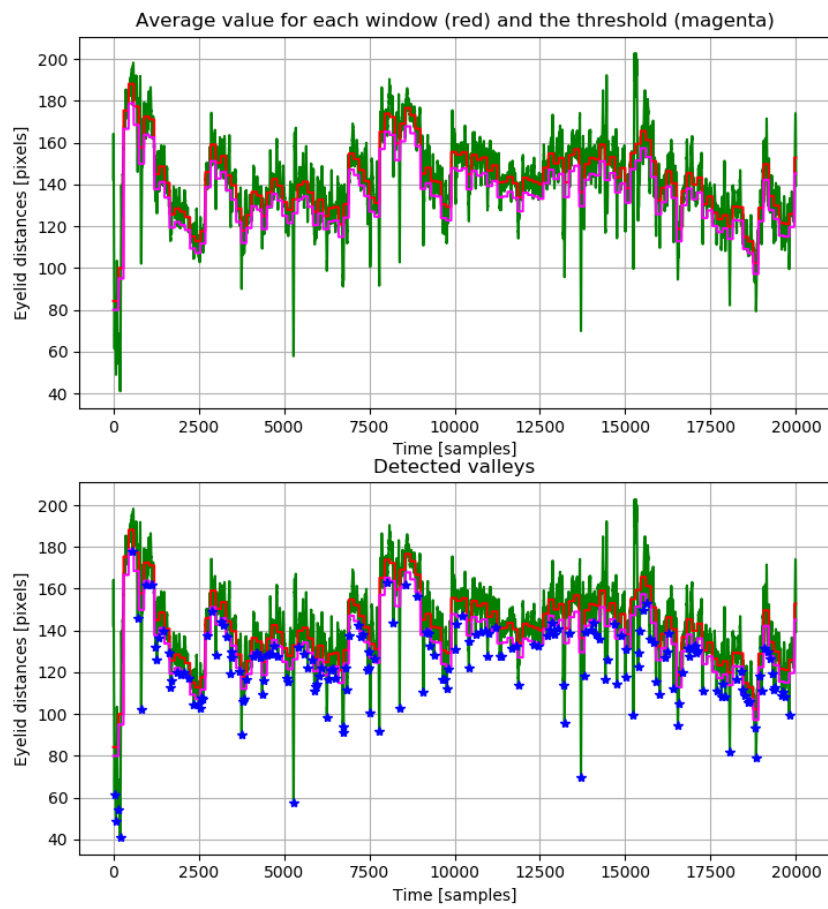
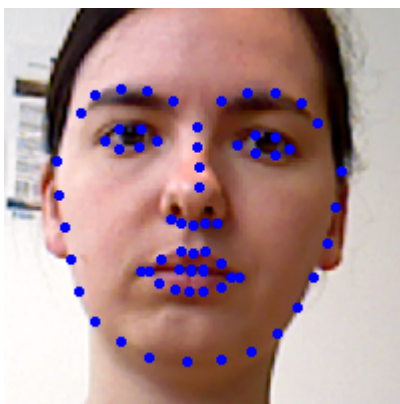


FIGURE 4.6: Offline blink detection steps: (top) Average values for each window and the threshold; (bottom) Detected valleys



(A)



(B)

FIGURE 4.7: (A) Facial feature points; (B) Eyes feature points (Agrigoroaie et al., 2017a)

the participant looks either away or towards the ground. To reliably detect blinks, the participants have to look directly at the camera.

Blinks are highly subjective (Nourbakhsh et al., 2013). Therefore, when using them to differentiate between certain tasks, it is important to calibrate them for each individual. The calibration is performed by dividing the value of the parameter during interaction j by the mean value of all the same parameters over all interactions of individual i (see Equation (4.1) taken from (Nourbakhsh et al., 2013). In Equation 4.1, m represents the total number of interactions performed by the subject.

$$\text{calibrated_feature}(i, j) = \frac{\text{feature}(i, j)}{\frac{1}{m} \sum_{j=1}^m \text{feature}(i, j)} \quad (4.1)$$

In this thesis, we considered the following parameters: the total number of blinks during each interaction, and the average blinking rate (AvgBR) (total number of blinks divided by the total interaction time). Both parameters were calibrated in order to eliminate subject dependency.

4.4 GSR

In Chapter 2, it was shown that the galvanic skin response (GSR) can be used to determine the arousal level and the emotional state of an individual, the cognitive load, as well as to differentiate between stress and cognitive load.

The GSR sensor used in this thesis was presented in Chapter 3. The sensor measures the skin resistance, having as output an analog voltage. First, the data has to be converted to skin resistance using the formula from the manufacturer's website (see Equation 4.2). The sensor reading values range between 0 and 1023. For recording the data, we used a frequency of 294 Hz. In order to remove a glitch in the raw signal, the manufacturer recommends the averaging of every 60 measurements. Then, the data can be converted to conductance ($\text{conductance} = \frac{1}{\text{resistance}}$) and used for the analysis.

$$\text{skin}_R = \frac{(1024 + 2 * \text{sensor_reading}) * 1000}{512 - \text{sensor_reading}} \quad (4.2)$$

A typical recorded GSR signal is shown in Figure 4.8. There are two types of analysis that can be performed on the GSR data: event based analysis and analysis of the entire interaction.

For the entire interaction analysis the following parameters are usually considered: accumulative GSR (AccGSR, i.e., the sum of the GSR values over the task time (Nourbakhsh et al., 2013)), the total number of peaks (only the peaks that were at least 2% of the total range of values were extracted) (Nourbakhsh et al., 2017), and the average GSR value (Shi et al., 2007). While, for the event based analysis there are four parameters of interest (Setz et al., 2010): latency time, rise time, amplitude, and recovery time (see Figure 4.9).

The **amplitude** represents the difference between the maximum value of the signal and the level at the event time. The **latency time** represents the time needed by the signal to increase with at least 10% compared to the level at event time. The **rise time** represents the time it takes the signal to increase from the value at the end of latency time up to the time when the maximum value is reached. The **recovery time** is computed as the time difference between the time when the signal reaches a level of 63% of the amplitude and the time when the maximum value is reached. The three times are measured in seconds.

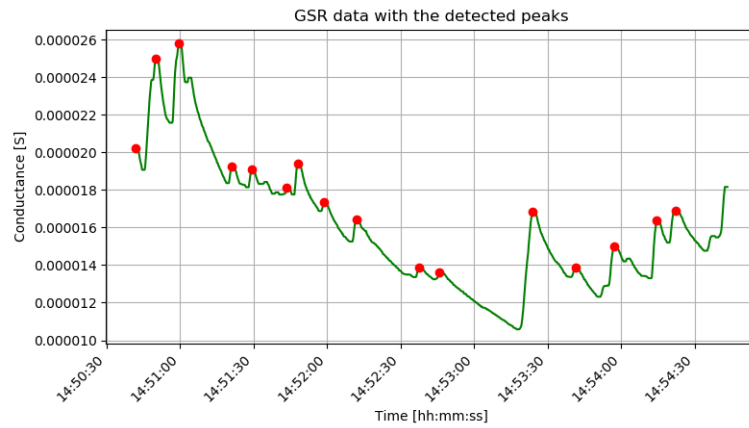


FIGURE 4.8: Filtered GSR signal with the detected peaks

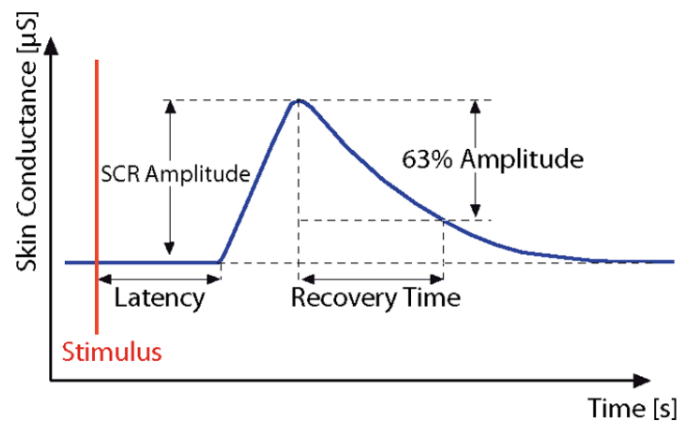


FIGURE 4.9: Ideal GSR signal with the computed features (Setz et al., 2010)

As blinks, GSR parameters are also highly subjective (Nourbakhsh et al., 2013). Therefore, when using them to differentiate between certain tasks, it is important to calibrate them for each individual. The calibration is performed by dividing the value of the parameter during interaction j by the mean value of all the same parameters over all interactions of individual i (see Equation (4.1) taken from (Nourbakhsh et al., 2013)). In Equation 4.1, m represents the total number of interactions performed by the subject.

In the second part of this chapter we are going to present the psychological questionnaires and how each trait is defined and computed.

4.5 Questionnaires

For this thesis, we define the user profile of an individual by using the results from psychological questionnaires. We have used personality questionnaires (EPQ, RST-PQ), morningness-eveningness questionnaire, and sensory profile. Next, each questionnaire is briefly described, as well as how each trait is computed.

4.5.1 Eysenck Personality Questionnaire (EPQ)

EPQ (Eysenck et al., 1975) (see Annex D) was developed to measure four personality traits: extraversion (E), neuroticism (N), psychotism (P), and lie/social desirability (L). The model, which was developed by Eysenck, considers that these personality traits have an underlying physiological basis. The physiological basis for extraversion is cortical arousal (Eysenck, 1983), which can be measured by skin conductance or EEG. The model developed by Eysenck states that introverted individuals tend to have higher levels of arousal in non-stressing conditions (Stelmack, 1981), as a result their electrodermal activity (EDA) is greater than that of extraverted individuals. The activation of the limbic system is related to neuroticism (Eysenck, 1983). In (Eysenck, 1967), Eysenck associated neuroticism to the emotionality level. Psychotism is characterized by hostile, suspicious, impersonal, and aggressive personality (Eysenck et al., 1976). It was shown that males tend to have much higher psychotism scores than females (Eysenck, 1983). As a result, testosterone levels could be used as an indicator of psychotism.

Research confirms that the personality traits are determined by biological factors (Eysenck, 1991). Each of the personality traits are measured on a scale from 0 to 12. Results can be categorized as low or high by using a threshold equal to 6 or to the median value in the chosen population.

4.5.2 The Reinforcement Sensitivity Theory Personality Questionnaire (RST-PQ)

RST-PQ (Corr et al., 2016) (see Annex E) was developed by Corr and is based on the theoretical analysis of the Reinforcement Sensitivity Theory (RST) proposed by Gray (Gray et al., 2000). RST is a neuroscience theory of personality, which proposes that the personality of a person is based on three neurobehavioral systems, which are responsible for appetitive and aversive motivation. RST-PQ measures the Behavior Activation System (BAS) (system responsible for the reactions towards all rewarding - appetitive - stimuli), the Fight, Flight, Freeze System (FFFS) (system responsible for the reactions towards aversive stimuli), and the Behavior Inhibition System (BIS) (system responsible for the reactions in the situations where there is conflict between approach (BAS) and avoidance (FFFS) situations). For the BAS there are 4 subscales:

Reward Interest (BAS-RI), Reward Reactivity (BAS-RR), Goal-Driven Persistence (BAS-GDP), and Impulsivity (BAS-I). As the Eysenck model, which was the starting point of RST, RST relates all personality traits to underlying physiological basis. Each system is measured on a scale from 1 to 4. Each participant's results were categorized as either low or high by using a threshold equal to 2 or the median value in the chosen population.

4.5.3 Morningness-Eveningness Questionnaire (MEQ)

MEQ (Horne et al., 1976) (see Annex B) was the first self-assessment questionnaire that was developed to determine morningness-eveningness in human circadian rhythm. Since the 1970's more and more attention has been given to an individual's preference for the optimum time of the day when certain activities (e.g., physical, cognitive) are carried out. The questionnaire consists of 19 questions and provides a score in the range 16 to 86. Based on this score an individual can be of one of five types: definite evening (scores between 16 and 30), moderate evening (scores between 31 and 41), intermediate (scores between 42 and 58), moderate morning (scores between 59 and 69), and definite morning (scores between 70 and 86). Research has shown that depending on the morningness-eveningness type, individuals might have better cognitive performance at different times of the day. It was shown (Horne et al., 1976) that evening individuals perform better cognitively in later hours of the day, than morning individuals. This questionnaire was used in order to understand whether being a morning type or an evening type can affect the cognitive performance. In (Randler, 2008) it was shown that morning type individuals show higher cognitive and physical performance in the morning hours, while evening individuals perform better during late hours.

4.5.4 Adult/Adolescent Sensory Profile Questionnaire (AASP)

AASP (Brown et al., 2002) was developed to measure the sensory profile of an individual. It is based on the four-quadrant model developed by Winnie Dunn (Dunn, 1997), which measures responses to everyday sensory experiences. The sensory profile was initially proposed for children, but research showed that it can be applied to adults as well (Brown et al., 2001). The model proposes that there is a relationship between the neurological threshold (low and high), and the behavioral response (accordance and counteract). The four-quadrants are: Low Registration (AASP-LR) - passive behavior and high neurological threshold, Sensation Seeking (AASP-SS) - active behavior and high neurological threshold, Sensation Avoiding (AASP-SA) - active behavior and low neurological threshold, and Sensory Sensitivity (AASP-Sens) - passive behavior and low neurological behavior. Each of the four quadrants is characterized by six sensory processing categories (i.e., taste/smell, movement, visual, touch, activity level, auditory). In this research, we are only interested in the visual and auditory processing categories. The sensory processing categories are measured on a scale from 1 to 5. The results of the participants were categorized as either low (a score less than 2.5) or high (a score higher than or equal to 2.5).

4.6 Conclusion and Contribution

In the first part of this chapter, we have presented how different physiological parameters are detected, extracted and analysed. More specifically, we have presented how the thermal, RGB and GSR data are used for this thesis. In the second part of the

chapter, we have presented four psychological questionnaires (EPQ, RST-PQ, MEQ and AASP), their respective traits and how they are computed. In the next chapter, we are going to present a study in which we investigate the relationship between cognitive performance and physiological response.

My contribution is represented by the development of a face detector for thermal images, as well as a thermal feature point predictor for 11 key feature points on the face. Based on these feature points multiple regions of interest can be defined. Furthermore, an algorithm to detect blinks based on the eyelid distance was developed and a software to extract and analyse GSR parameters was also written.

A part of the work described in this chapter was published in “Contactless Physiological Data Analysis for User Quality of Life Improving by Using a Humanoid Social Robot”, in the 19th International Conference on Image Analysis and Processing, 2017 (Agrigoroaie et al., 2017a).

Chapter 5

RQ1: Relationship between Cognitive Performance and Physiological response

5.1	Introduction	33
5.2	Experimental design	34
5.2.1	Robotic Platform and Sensors	34
5.2.2	Questionnaires	34
5.2.3	Scenario	35
5.2.4	Conditions	35
5.2.5	Participants	36
5.2.6	Interaction time	36
5.3	Results	37
5.3.1	Task performance based on condition parameters	38
5.3.2	Task performance based on user profile	39
5.3.3	Physiological parameters variation based on condition parameters	42
5.3.4	Physiological parameters variation based on user profile	43
5.3.5	Results for morning individuals	44
5.3.6	Results for evening individuals	45
5.3.7	Correlation results	47
5.3.8	Other results	48
5.4	Classification results	48
5.5	Discussion	51
5.6	Conclusions	52
5.7	Contribution	52

5.1 Introduction

We have shown in Chapter 2 that interfaces designed for HRI should lower the cognitive load on the user (Yanco et al., 2004). In order to do this, first, the cognitive load should be detected. But, how can a robot determine the level of cognitive load? In (Kramer, 1991), Kramer reviews the following physiological parameters: the electroencephalographic activity (EEG), pupil diameter, endogenous eye blinks, event-related brain potentials, electrodermal activity (EDA), and cardiac activity. More recently, the authors of (Charles et al., 2019) have performed a literature review of the measures used for mental workload. They investigate previous works related to electrocardiac activity, blood pressure, ocular measures, respiration, skin based measures, and brain measures. These changes are dependent on multiple aspects: user profile, type of task, time of the day, mood, etc.

Another important aspect to be considered is represented by individuals' preference for the optimum time of the day when certain activities are carried out (Adan, 1991). Research has shown that cognitive performance is dependent on the morningness-eveningness type (ME-type): morning type individuals perform better cognitively in the morning hours than in the evening hours (Horne et al., 1976; Adan, 1991; Adan et al., 2012).

In this chapter, we define the user profile as a function of personality (Eysenck et al., 1975; Corr et al., 2016), sensory profile (Brown et al., 2002) and ME-type (Horne et al., 1976).

The purpose of this chapter is to investigate the role played by the morningness-eveningness type in the cognitive performance as well as to investigate how the different physiological parameters vary during a cognitive task. Therefore, we defined two research questions to be investigated:

1. **Can the ME-type of an individual influence the performance on a cognitive task with a robot?**
2. **What is the importance of different interaction styles (i.e., encouraging, and stressing) of a humanoid robot in both the performance during the task and in the variation of the physiological parameters?**

The cognitive task that we chose was successfully used before as a laboratory stressor (i.e., Stroop task (Stroop, 1935), see also Chapter 3.5.1). The task was performed by each participant three times: first, without the robot, then in the presence of an encouraging robot, and lastly in the presence of a stressing robot. The order of the two tasks performed in the presence of the robot was randomized between participants. The role of personality and sensory profile was also investigated.

Our results show that there is a slight interaction between the time of the day when a task is performed, the ME-type and the task performance. Furthermore, in relation to the sensory profile, we have found that morning type individuals need low intensity stimuli, while evening type individuals require high intensity stimuli.

Next, we are going to present the experimental design for this study, the results, and a discussion of the obtained results.

5.2 Experimental design

5.2.1 Robotic Platform and Sensors

For this study, the TIAGo robot (Pages et al., 2016) (also presented in Chapter 3.2) was used. The physiological data was recorded using the three sensors presented in Chapter 3.4: RGB-D camera, a thermal camera, and a GSR sensor. The two cameras were positioned on the same table where the task was performed (see Figure 5.1). The GSR sensor was positioned on the non-dominant hand of the participants, on the ring and middle fingers.

5.2.2 Questionnaires

Each participant filled out the following questionnaires (see also Chapter 4.5): EPQ (for personality), MEQ (for ME-type), RST-PQ (for personality), AASP (for sensory profile). Moreover, each participant filled a post experiment questionnaire.

A custom designed post-questionnaire with 20 questions was developed in order to assess how the participants perceived the task, the behavior of the robot, and



FIGURE 5.1: Experimental setup

the different stimuli used to throughout the interaction. The questionnaire was filled by the participants after each behavior of the robot. Some questions were of Yes/No type (e.g., “Was the speech of the robot too loud?”, “Did you find the instructions clear?”), while others were open ended questions (e.g., “Describe how the behavior of the robot affected your performance”, “Describe how the presence of the timer affected your performance”). On a 5-point Likert scale from 1 (“Very easy”) to 5 (“Very hard”), we asked the participants to rate the difficulty of the task. Without divulging the type of behavior displayed by the robot, we asked the participants to choose between three possible behaviors (i.e., Neutral, Encouraging, or Stressing). And lastly, on a scale from 1 (“I disagree completely”) to 5 (“I agree completely”), the participants had to rate if the robot was stressing, and if the robot was encouraging.

5.2.3 Scenario

For this study, each participant was greeted by the experimenter who explained the experimental task. Upon making sure that the participants knew what they had to do, they were asked to fill in the questionnaires presented in Section 5.2.2.

The experimental task chosen for this study is a variation of the original word-color Stroop test (Stroop, 1935), also presented in Chapter 3.5.1. Each participant was seated at a table, on top of which was only a monitor, a speaker, the RGB-D and thermal cameras, and a mouse (as shown in Figure 5.1). They had to use the mouse to press on the button corresponding to the color of the text written on the screen. We used the Stroop task without any time limit, therefore, only the reaction time (RT) for each trial was recorded. The speaker was used to either play or not a timer sound. For each condition the same number of congruent and incongruent trials were shown in a random order.

5.2.4 Conditions

This experiment was designed around three conditions: control - no robot, Encouraging Robot, and Stressing Robot. Each participant performed all three conditions. The first condition was always the control condition, while the order of the encouraging and stressing conditions was randomized between participants. After each condition there was a 2 minutes pause with some relaxing classical music. The movement of

the robot was remotely controlled by the experimenter.

Condition 1: Control

The purpose of the Control condition was for the participants to familiarize themselves with the task. This is an important phase as we wanted to counteract the novelty factor that would influence the physiological parameters. This condition consisted of 60 trials. The last 20 trials were considered as the control for the RT for each participant.

Condition 2. Encouraging Robot

This condition consisted of 50 trials. The robot maintained a fixed position and displayed only auditory stimuli (encouraging verbal content, medium volume for robot speech, and no sound for the timer). Every 4 seconds, the robot selected in a random order an encouraging verbal content from a predefined list of phrases (e.g., “Don’t look at the time”, “Relax”, “Think carefully what you have to do”). No matter how the participant performed, an encouraging content was given by the robot. The volume for the speech was set at 50% of the maximum volume of the robot speakers.

Condition 3. Stressing Robot

As with the Encouraging Robot condition, this condition consisted of 50 trials. The robot was controlled to continuously move on the left side of the participants. It also turned around in order to distract the participants from performing the task. The robot displayed a mixture of both visual (movement of the robot) and auditory (stressing verbal content, high volume for robot speech, sound for the timer) stimuli. The stressing verbal phrases were randomly selected every 4 seconds from a given list of phrases (e.g., “Faster, faster”, “Hurry up”, “Look! The time is ticking”) and given to the participants. The stressing content was given no matter how good or bad the participant performed. The volume for the speech was set at 80% of the maximum volume of the robot speakers.

During each condition, the same number of congruent and incongruent trials were given in a random order.

5.2.5 Participants

A total of 24 participants (5 females and 19 males with an average age of 27.38, $SD=6.02$) agreed to take part in this experiment. All participants have a technical background. Regarding the participants knowledge about robotics, on a scale from 1 ("Not at all") to 5 ("Very much"), 5 participants reported little knowledge, 8 knew somewhat, while 11 knew much or very much.

Table 5.1 presents the distribution of the participants based on their results from the questionnaires presented in Section 5.2.2. Considering the distribution of the participants, the following questionnaire results are used for analysis: E and N (from EPQ), the ME-type, BAS-I from RST-PQ, and AASP-Sens both visual and auditory, AASP-SS visual, AASP-SA auditory, and AASP-LR auditory.

5.2.6 Interaction time

As for this study, we are investigating the effect of ME-type on task performance, it is important to consider at what time of the day the interactions took place. The earliest interaction started at 09h47 (and finished at 10h05), while the latest one started at 18h54 (and finished at 19h13). When considering the interaction time, we

TABLE 5.1: Participants distribution based on questionnaires results

MEQ		EPQ			
morning type	evening type	Category	E	P	N
11	13	low	10	21	11
		high	14	3	13

AASP				
Category	low visual	high visual	low auditory	high auditory
AASP-LR	21	3	13	11
AASP-SS	12	12	8	16
AASP-SA	18	6	13	11
AASP-Sens	14	10	14	10

RST						
Category	FFFS	BIS	BAS I	BAS RI	BAS RR	BAS GDP
low	16	8	13	2	7	2
high	8	16	11	22	17	22

considered the hour at which the interaction finished. All interactions that finished before 15h00 were considered as morning time interactions, while the interactions that finished after 15h00 were considered as evening time interactions. In Figure 5.2 we show the distribution of the participants by taking into consideration their ME-type and the hour at which they performed the experimental task. For example, between 17h00 and 17h59, four participants did the experiment. Out of the four, three are of morning type (depicted with a sun) and one is of evening type (depicted with a crescent).

5.3 Results

Taking into consideration the task performed by the participants and the measures recorded, the results are divided into multiple parts: task performance based on condition parameters, task performance based on user profile, physiological parameters variation based on condition parameters, physiological parameters variation based on user profile, results for morning type individuals, results for the evening type individuals, and analysis of the correlation between the dependent variables.

The task performance was measured by the average reaction time (RT) during the congruent trials (RTCong), average RT during incongruent trials (RTIncong), and the average RT during all trials (RTTotal). The condition parameters considered are the condition (Control, Encouraging Robot, and Stressing Robot) and the interaction time (morning or evening).

The physiological parameters considered for the analysis are: the AccGSR, the number of peaks, the average blinking rate (AvgBR), the total number of blinks, the rate of change of the facial temperature in the nose, forehead, and perinasal regions.

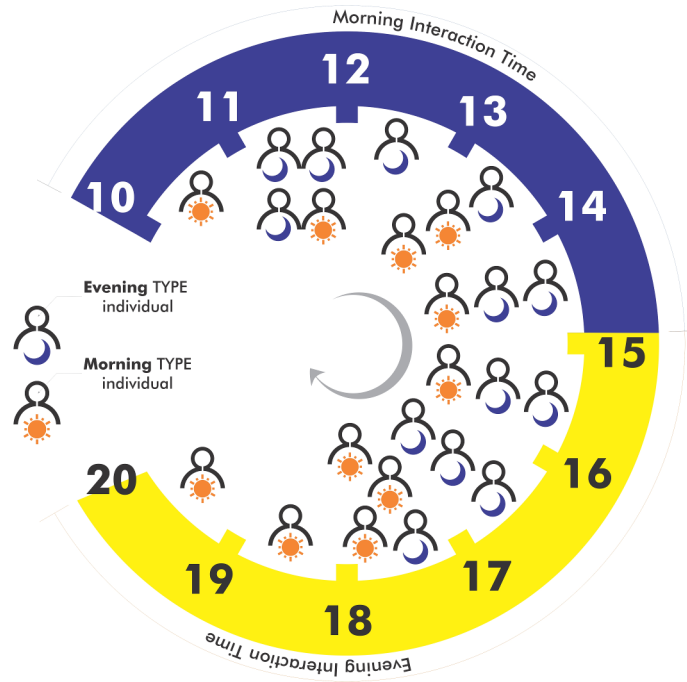


FIGURE 5.2: Participants distribution based on ME-type and the time of the day when the task was performed

For the analysis based on the condition parameters, we used the calibrated GSR and blink parameters, while for the other analyses, we used the uncalibrated parameters, as we wanted to find the subjective differences.

5.3.1 Task performance based on condition parameters

First, we checked if our dependent variables are normally distributed. For this purpose, we applied a normality test (i.e., Shapiro-Wilk). None of the three types of RT are normally distributed ($W = 0.8, p = 2.59e-8$; $W = 0.93, p = 0.001$; $W = 0.90, p = 7e-05$, respectively). Therefore, the Kruskal-Wallis non-parametric statistical test was applied.

A significant result was found for the RT during all trials (RTTotal) based on the condition ($\chi^2 = 7.91, p = 0.019$). A pairwise comparison using Wilcoxon rank sum test with a Bonferroni correction revealed significant differences between the Control and Stressing Robot conditions ($p = 0.032$). During the Control condition ($M_{RTTotal} = 1.48, SD = 0.38$) the participants had a higher RTTotal than during the Encouraging Robot condition ($M_{RTTotal} = 1.25, SD = 0.23$), and a significantly higher RTTotal than during the Stressing Robot condition ($M_{RTTotal} = 1.22, SD = 0.2$). The interaction time did not have any influence on the performance.

Discussion: for the first analysis that we performed, we found that when considering all trials, the participants of this study were slower during the last 20 trials of the Control condition than during the two robot conditions. One possible explanation for this result is that during the two robot conditions, at every four seconds there was either an encouraging or a stressing verbal content from the robot. We believe that

TABLE 5.2: Significant results for task performance based on user profile

Dependent variable	Factor	χ^2	p-value
RTCong	Extraversion	8.41	0.003
RTIncong	Extraversion	3.68	0.055
RTTotal	Extraversion	5.42	0.019
Dependent variable	Factor	Mean value	SD
RTCong	Introverted	1.09	0.27
	Extraverted	1.21	0.26
RTIncong	Introverted	1.34	0.41
	Extraverted	1.42	0.25
RTTotal	Introverted	1.24	0.32
	Extraverted	1.37	0.27

this gave the participants an incentive to be faster. Some of the answers from the post-questionnaire open question “Describe how the behavior of the robot affected your performance” prove this aspect: “He was encouraging and supportive” (Encouraging Robot), “Calm down” (Encouraging Robot), “Disturbing” (Encouraging Robot), “It was stressful, I wanted to finish faster so he could stop talking” (Stressing Robot), “It was stressing, but maybe this helped me to answer faster” (Stressing Robot), or “It somewhat made me lose focus, but it also motivated me to perform better” (Stressing Robot).

5.3.2 Task performance based on user profile

The user profile is composed of the following traits: ME-type, extraversion, neuroticism, impulsivity, and the sensory profile. The significant results after applying the Kruskal-Wallis non-parametric statistical test are shown in Table 5.2. The only significant factor for the performance is the extraversion personality trait, with introverted individuals performing better (i.e., faster) than extraverted individuals for all three types of RT (RTCong, RTIncong, and RTTotal). This result is more significant for the congruent trials (RTCong, $p = 0.003$). For the incongruent trials (RTIncong) the result only approaches significance ($p = 0.055$). The result for the congruent trials is graphically shown in Figure 5.3.

Moreover, in Table 5.2, we also show the mean and standard deviation values for introverted and extraverted individuals for all three types of RT.

Next, we investigated the interaction between the interaction time and the ME-type when considering the performance parameters (RTCong, RTIncong, and RTTotal). We performed this investigation as we believe that if such an interaction exists, this means that a social robot could adapt its behavior in order to increase user task performance. For example, the ENRICHME system was developed so that elderly individuals with MCI can maintain or, if possible, improve their cognitive abilities. Therefore, the robot can propose to the individual it interacts with to perform the cognitive games at the optimum time of the day depending on his/her ME-type.

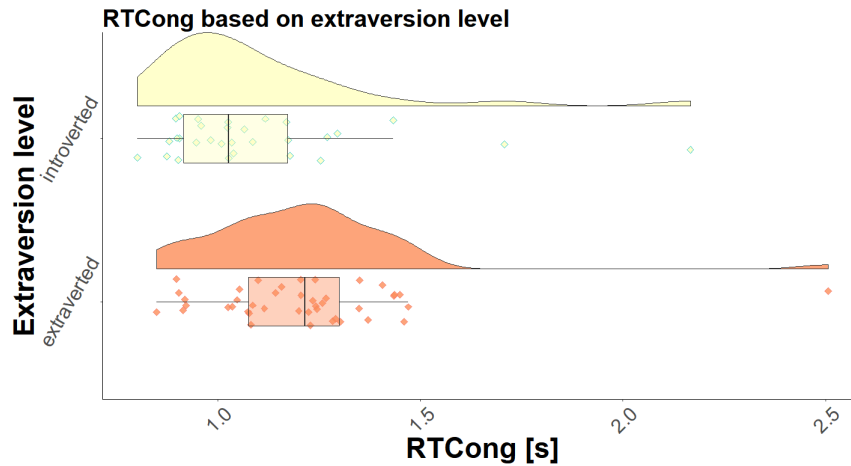
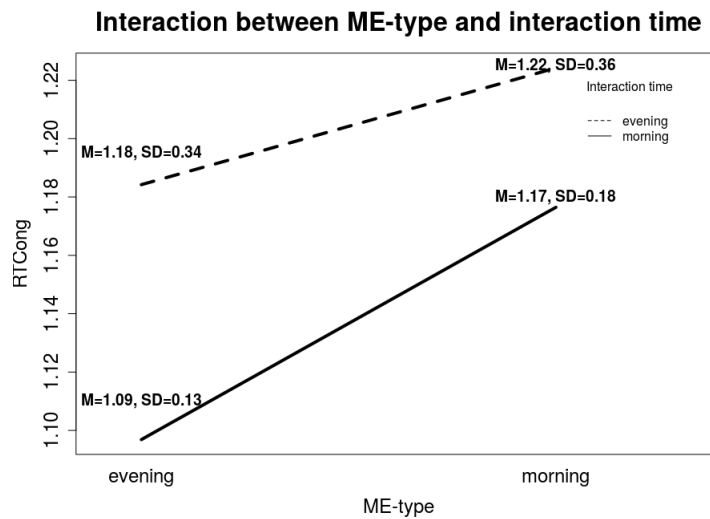


FIGURE 5.3: Reaction Time during congruent trials based on the extraversion trait of EPQ

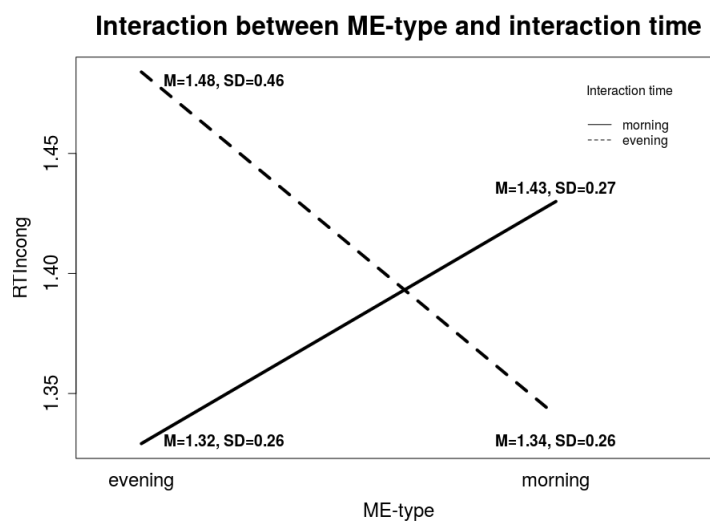
In Figures 5.4a, 5.4b, 5.4c are shown the interaction plots for the interaction time, ME-type and RTCong, RTIncong, and RTTotal, respectively. The mean and standard deviation values for each group are shown in the corresponding figures.

By analyzing the three plots, it can be seen that while for the easy task (i.e., congruent trials) there is no interaction between the interaction time and the ME-type (i.e., the two lines do not intersect), for the difficult task (i.e., incongruent trials), there is an interaction between the two (i.e., the two lines intersect). For the incongruent trials, morning type individuals performed better in the evening hours ($M = 1.34, SD = 0.26$) compared to the morning hours ($M = 1.43, SD = 0.27$), while evening type individuals performed better in the morning hours ($M = 1.32, SD = 0.26$) compared to the evening hours ($M = 1.48, SD = 0.46$). Even if the interactions are not significant, these results present important insights into the influence of ME-type and the interaction time on task performance depending on the task difficulty (i.e., congruent trials or incongruent trials).

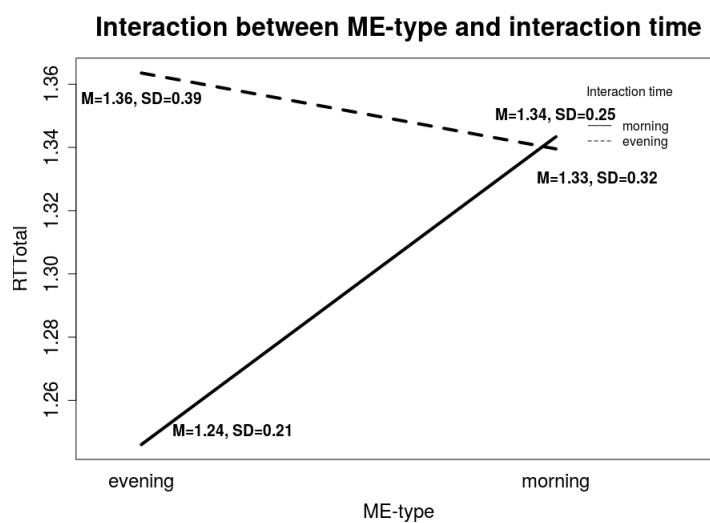
Discussion: While the interactions found are not significant, they still are of importance for HRI. They show that there are differences in how individuals perform depending on the time of the day, the task difficulty, and their ME-type. However, further studies are required in order to better understand these results and to investigate why our tendencies are different from the ones found in the literature (i.e., morning type individuals should perform better in the morning hours, while evening type individuals should perform better in the evening hours). Next, we provide some of the reasons for why we believe we obtained these results. The interaction times were considered as either morning (before 15h00) or evening (after 15h00). We should also consider the afternoon hours. Furthermore, we divided our participants into two groups (either morning or evening). However, according to (Adan, 1991), it is important to consider for analysis also the Intermediate type individuals. If divided into three groups (i.e., morning, intermediate, and evening), most of our participants would fit in the Intermediate type category (15 participants out of 24). Therefore, we consider that further investigation is required in order to better understand these results. Moreover, different tasks and distinct difficulty levels have to be considered as well.



(A) Congruent trials



(B) Incongruent trials



(C) All trials

FIGURE 5.4: Interaction plot for ME-type and the interaction time

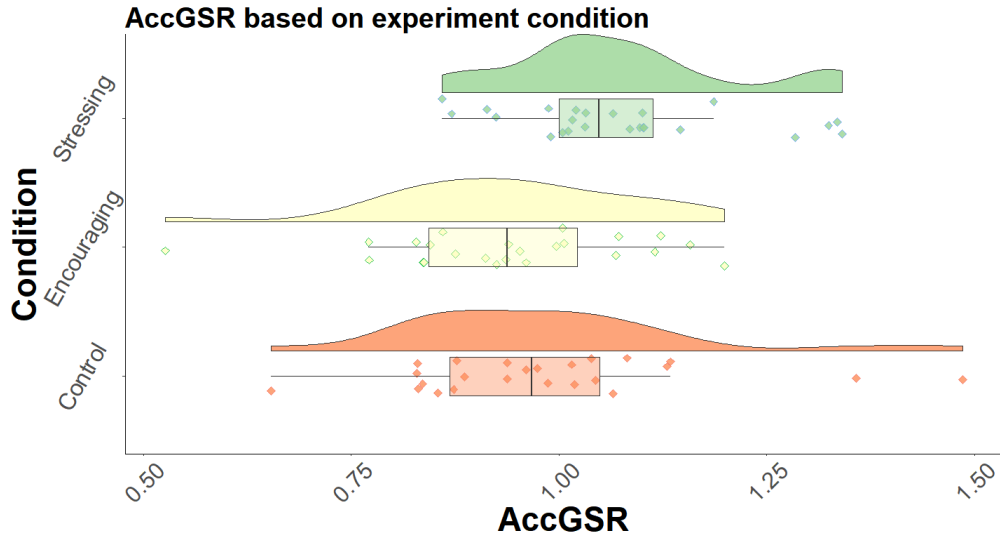


FIGURE 5.5: AccGSR based on condition

5.3.3 Physiological parameters variation based on condition parameters

The Shapiro-Wilk normality test showed that none of the temperature parameters are normally distributed, while the GSR and blink parameters are all normally distributed. As a result, an one-way Anova analysis was performed on the GSR and blink parameters, while Kruskal-Wallis non-parametric statistical test was applied on the temperature parameters.

For the condition factor significant results were found for the following parameters: AccGSR ($F(2, 69) = 4.9, p = 0.01$), the total number of peaks ($F(2, 69) = 16.22, p = 1.68e-07$), and the AvgBR ($F(2, 69) = 7.64, p = 0.001$).

For the AccGSR (see Figure 5.5), a pairwise t-test with a Bonferroni correction revealed significant differences between the Stressing and Encouraging Robot conditions. During the Stressing Robot condition, the AccGSR was significantly higher than during the Encouraging Robot condition ($t = 3.08, p = 0.0089$).

For the total number of peaks significant differences were found between the Control and the Robot conditions (Encouraging Robot ($t = -4.00, p = 0.0004$), Stressing Robot ($t = -5.5, p < 0.0001$) conditions, respectively). During the Control condition there were significantly more GSR peaks than during the Robot conditions.

Contrary to the number of GSR peaks, for the blinking rate (AvgBR) significantly lower values were found for the Control condition compared to the Robot conditions (Encouraging Robot ($t = 3.63, p = 0.0016$), and Stressing Robot ($t = 3.06, p = 0.0094$) conditions, respectively).

For the interaction time, significant results were found for the rate of change for the temperature in the nose ($\chi^2 = 4.14, p = 0.04$) and the perinasal ($\chi^2 = 5.19, p = 0.02$) regions. For both regions, the rate of change for the facial temperature is greater in the morning hours than in the evening hours.

Discussion: The results for the AccGSR are in accordance with the literature. Higher cognitive loads are associated with higher values for the GSR parameters (Nourbakhsh et al., 2013; Nourbakhsh et al., 2017). During the Stressing Robot condition our participants had significantly higher AccGSR values than during the Encouraging Robot condition. When considering the blinking results, we can interpret these results in

TABLE 5.3: Significant results for physiological parameters variation based on user profile

Dependent variable	Factor	χ^2	p-value
Slope forehead	AASP-Sens auditory	3.81	0.050
Slope forehead	AASP-LR auditory	3.95	0.046
Slope nose	AASP-SA auditory	9.15	0.002
Slope perinasal	AASP-SA auditory	7.16	0.007

terms of the visual and auditory stimuli used by the robot. During the Control condition, the participants had only to perform the task, while in the two Robot conditions there were other stimuli that they had to pay attention to. The literature review showed that the exposure to visual stimuli decreases the blinking frequency. However, in our case the participants were exposed to both auditory and visual stimuli. Based on our condition descriptions, for the Encouraging Robot condition, we only used auditory stimuli, while during the Stressing Robot condition there was a mixture of auditory and visual stimuli. As also shown in Chapter 2, exposure to visual stimuli leads to a decrease in the frequency of the blinking rate, we would expect that during the Stressing Robot condition there should be a lower number of blinks. Our results show a mean value of $M = 1.11$ blinks/min and a standard deviation of $SD = 0.31$, during the Encouraging Robot condition compared to a mean value of $M = 1.06$ blinks/min and a standard deviation of $SD = 0.26$ during the Stressing Robot condition, and a mean value of $M = 0.81$ blinks/min with an $SD = 0.29$ for the Control condition. While the difference is not significant, we can see that during the Stressing Robot condition, the participants had a lower blinking rate than during the Encouraging Robot condition.

5.3.4 Physiological parameters variation based on user profile

The significant results are shown in Table 5.3.

For the temperature variation in the forehead region, for both factors (auditory AASP-Sens, and auditory AASP-LR, respectively) individuals with low scores showed a greater rate of change than individuals with high scores. For the auditory AASP-SA factor, the rate of change for the temperature in the nose, and perinasal regions, individuals with high scores showed a greater increase than individuals with low scores.

For the blink and GSR features, for the condition parameters analysis we considered the calibrated features (that eliminate the subjective differences). However, as in this situation we are interested in investigating the differences that exist between individuals, therefore we considered the uncalibrated features. Significant results were found for the number of blinks ($\chi^2 = 7.92, p = 0.0048$) and the average blink rate ($\chi^2 = 7.60, p = 0.0058$) depending on the ME-type. Evening type individuals blinked significantly less than morning type individuals.

Discussion: Our results show that the temperature variation in the three face ROIs (i.e., forehead, nose, perinasal region) is dependent on the sensory profile, while the two blink parameters (i.e., number of blinks, average blink rate) are dependent on the ME-type. We are not entirely sure why this happens. To the best of our knowledge the relationship between the sensory profile and the facial temperature variation, as

TABLE 5.4: Significant results for morning individuals

Dependent variable	Factor	χ^2 / F	p-value
Condition parameters			
AccGSR	condition	5.53	0.009
Slope nose	interaction time	4.94	0.026
Slope perinasal	interaction time	3.95	0.046
User profile			
RTCong	AASP-Sens auditory	3.95	0.046
RTIncong	AASP-Sens visual	0.40	0.043
RTTotal	AASP-Sens auditory	3.81	0.05
Slope perinasal	AASP-SA auditory	4.09	0.04
Slope nose	AASP-SA auditory	5.39	0.02
Slope forehead	BAS I	5.04	0.024
AccGSR	Extraversion	6.83	0.0089
AccGSR	AASP-SA auditory	5.52	0.018

well as that of the number of blinks and the ME-type has not been investigated. Therefore, we are not sure how to interpret these results. This will have to be further investigated.

5.3.5 Results for morning individuals

We performed the analysis based on the condition parameters (condition and interaction time) and based on the user profile. The significant results are summarized in Table 5.4.

For the AccGSR, a pairwise comparisons using t-tests with a Bonferroni correction revealed significant differences between the Control and the Stressing Robot conditions ($t = -0.38, p = 0.036$), and the Stressing and Encouraging Robot conditions ($t = 3.05, p = 0.014$). For the Stressing Robot conditions, significantly higher levels of AccGSR were found compared to the Encouraging Robot and the Control conditions. This result is also shown in Figure 5.6.

When considering the interaction time as the independent variable, we found that both for the nose region, and for the perinasal region, the variation of the temperature was higher during the morning hours than during the evening hours.

Regarding the user profile, for the morning type individuals, the sensory profile, the impulsivity trait of RST and the Extraversion trait of EPQ proved to be significant factors when looking at the performance, AccGSR, and the facial temperature variation.

Morning type introverted individuals have a significantly lower AccGSR than extroverted individuals.

Individuals with high scores for the auditory sensory processing category of AASP-Sens had a better performance for both RTCong ($p = 0.046$) and RTTotal ($p = 0.05$) than individuals with low scores. The opposite was found for the AASP-Sens visual processing category, with individuals with a low score performing significantly faster than individuals with high scores.

For both the nose and the perinasal regions, for the individuals with high scores for auditory AASP-SA the facial temperature increased, while for the individuals with

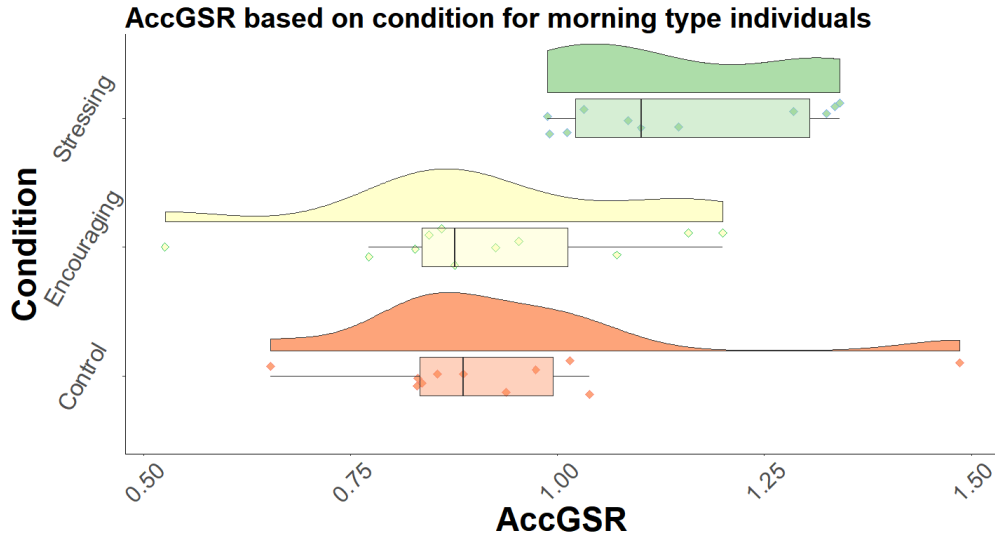


FIGURE 5.6: AccGSR for morning individuals based on condition

low scores, the temperature decreased. Regarding the variation of the temperature in the forehead region, for all individuals there was an increase in temperature, with a significantly higher increase for the individuals with low BAS_I scores compared to the individuals with high BAS_I scores.

Discussion: One of the final goals of this study is to find out how a social robot can use the ME-type of the individual it interacts with in order to better adapt its behavior. For that, it is important to investigate how morning and evening type individuals differ in performance and in the variation of their physiological parameters. The results presented in this section show us that it is important to consider both the sensory profile, the impulsivity and the extraversion personality trait when investigating the task performance and the facial temperature variation.

5.3.6 Results for evening individuals

For the evening type individuals, significant results were found for the peaks, AvgBR and the temperature variation in the nose region when considering the condition as factor.

For the peaks, a pairwise t-test using Bonferroni correction showed significant differences between all conditions (Control - Encouraging Robot $t = -3.81, p = 0.001$, Control - Stressing Robot $t = -6.27, p < 0.0001$, Encouraging Robot - Stressing Robot $t = -2.46, p = 0.055$). The highest number of peaks were found during the Control condition, followed by the Encouraging Robot condition, and lastly by the Stressing Robot condition.

For the temperature variation in the nose region, a Wilcoxon rank sum test with a Bonferroni correction revealed significant differences between the Encouraging and the Stressing Robot conditions ($p = 0.016$). During both robot conditions the facial temperature increased, with a higher rate for the Encouraging Robot.

When considering the AvgBR as a factor, a pairwise t-test using Bonferroni correction showed that during the Control condition the participants had a significantly lower blink rate than during the Encouraging Robot condition ($t = 3.14, p = 0.009$). No differences were found between the two robot conditions.

TABLE 5.5: Significant results for evening individuals

Dependent variable	Factor	χ^2 / F	p-value
Condition parameters			
Peaks	condition	20.01	< 0.0001
AvgBR	condition	5.34	0.009
Slope nose	condition	7.08	0.028
User profile			
RTCong	intro	5.07	0.024
RTCong	AASP-LR auditory	5.33	0.024
RTIncong	AASP-SS visual	3.88	0.04
RTIncong	AASP-LR auditory	4.95	0.02
RTTotal	AASP-SS visual	5.6	0.017
RTTotal	AASP-LR auditory	4.00	0.045
Slope nose	neuro	5.96	0.014
Slope forehead	AASP-Sens auditory	3.91	0.047

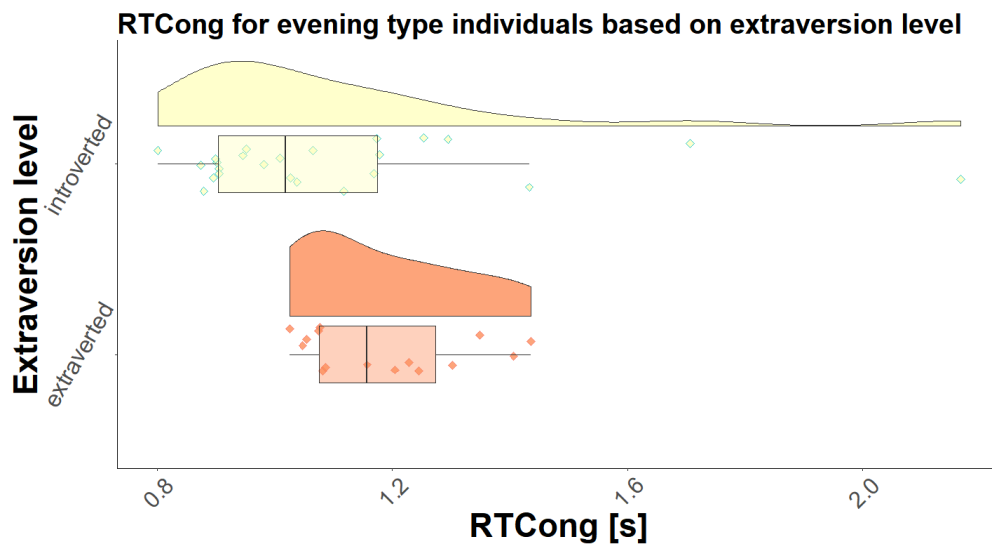


FIGURE 5.7: RTCong for evening individuals based on the personality trait of extraversion

TABLE 5.6: Significant results for correlations

Dependent variable I	Dependent variable II	rho	p-value
RTCong	RTIncong	0.71	3.2e-12
RTCong	RTTotal	0.86	< 2.2e-16
RTIncong	RTTotal	0.88	< 2.2e-16
AccGSR	peaks	0.57	1.3e-07
Blinks	AvgBR	0.95	< 2.2e-16
Slope forehead	peaks	-0.25	0.034
Slope forehead	Slope nose	0.53	4.6e-06
Slope forehead	Slope perinasal	0.5	1.2e-05
Slope nose	peaks	-0.24	0.047
Slope nose	AccGSR	-0.25	0.04
Slope nose	Slope perinasal	0.87	< 2.2e-16

Introverted evening type individuals performed significantly better than extraverted individuals ($p = 0.024$, see also Figure 5.7).

Individuals with high scores for visual AASP-SS performed better than individuals with low scores when considering the performance parameters, RTIncong - $p = 0.04$, and RTTotal - $p = 0.017$.

However, the opposite was found for auditory AASP-LR, with better performance found for individuals with low scores (RTCong - $p = 0.024$, RTIncong - $p = 0.02$, RTTotal - $p = 0.045$, respectively).

For the individuals with low neuroticism levels the temperature in the nose region decreased while for the individuals with high scores the temperature increased ($p = 0.014$).

Discussion: The results from the morning type individuals and those presented in this section show that, when taking into consideration the sensory profile of the participants, the dependent variables of the morning type individuals were mostly influenced by the Sensation Avoiding and the Sensory Sensitivity quadrants, while the dependent variables of the evening type individuals were mostly influenced by the Sensation Seeking and Low Registration quadrants. Both Sensation Avoiding and Sensory Sensitivity are characterized by low neurological threshold (i.e., low-intensity stimulus is sufficient for an individual to react). On the contrary, Low Registration and Sensation Seeking quadrants are both characterized by a high neurological threshold (i.e., individuals require a high-intensity stimulus or a longer time to react to the same stimulus). This represents a very interesting result which could enable us to better understand the type and strength of stimulus needed by each ME-type individual.

5.3.7 Correlation results

Lastly, we investigated the correlation that exists between the performance parameters and the variation of the physiological parameters; and the correlation between the physiological parameters. For the normally distributed parameters, we applied the Pearson correlation (Sprent et al., 2000), while for the non-normally distributed

parameters we applied Spearman’s rank correlation (Sprent et al., 2000). The significant results are shown in Table 5.6.

Significant positive correlation was found between all performance parameters. Furthermore, positive correlations were found for the GSR parameters (i.e., AccGSR, peaks) and the blink parameters (i.e., blinks and AvgBR). Positive correlations were found for the forehead, perinasal, and nose regions.

We also found a negative correlation between the temperature variation in the forehead region and the number of uncalibrated peaks, as well as the temperature variation in the nose area and the number of peaks and the AccGSR.

Discussion: As the values for AccGSR and the number of peaks increase, the temperature in the nose region decreases. It is already known from the literature that high values of AccGSR correspond to high cognitive load, while a decrease of temperature in the nose region is associated with high cognitive load as well (Abdelrahman et al., 2017).

5.3.8 Other results

We wanted to investigate how much does the previous knowledge about robotics influence the performance. We divided the participants into three groups: little (N=5), somewhat (N=8), and much or very much (N=11).

First, we performed the analysis on the performance parameters (RTCong, RT-Incong, RTTotal). A significant result was found only for RTCong ($\chi^2 = 10.35$, $p = 0.0056$). A pairwise comparison using Wilcoxon rank sum test with a Bonferroni correction revealed significant differences between the little and somewhat groups ($p = 0.01$), and little and much groups ($p = 0.0085$). In both cases, individuals in the little knowledge group performed significantly worse than the participants in either of the other two groups.

And lastly, we wanted to investigate how the participants perceived the robot while displaying the two behaviors. When the robot displayed an Encouraging behavior, out of the total 24 participants, half of them considered the robot encouraging, 9 considered it stressful, while 3 considered it neutral. On the other hand, 21 participants out of 24 thought that the Stressing robot was stressing, while 3 considered it as neutral.

5.4 Classification results

In (Nourbakhsh et al., 2013; Nourbakhsh et al., 2017), Nourbakhsh et al., have used Support Vector Machine (SVM) and Naïve Bayes classifiers for cognitive load classification. We wanted to investigate if our physiological parameters (i.e., AccGSR, peaks, blinks, blink rate, nose temperature variation, forehead temperature variation) could be used in a two-class classification problem (i.e., the two robot conditions) and a three-class classification problem (i.e., the three conditions: Control, Encouraging Robot, Stressing Robot). First, we standardized the features (i.e., we removed the mean and scaled them to unit variance). Each physiological parameter is characterized by one data point for each participant and for each condition. We used 70 % of the data for training, and 30 % of the data for testing. We performed a k-fold cross-validation (with k=3). This cross-validation ensures that each observation is used only once for testing. The final accuracy is the average accuracy from all rounds. In Table 5.7, we show the accuracy when training the classifiers with each physiological

TABLE 5.7: Classification accuracy for individual physiological parameters

2-Class Classification		
Parameter	Classification Algorithm	
	SVM [%]	Naïve Bayes [%]
AccGSR	70.8	70.8
GSR Peaks	52.1	33.3
Blinks	54.2	52.1
Blink rate	58.3	52.1
Forehead temp	60.4	52.1
Nose temp	56.3	54.2
3-Class Classification		
Parameter	Classification Algorithm	
	SVM [%]	Naïve Bayes [%]
AccGSR	25.0	34.7
GSR Peaks	51.4	47.2
Blinks	25.0	30.6
Blink rate	51.4	41.7
Forehead temp	38.9	29.2
Nose temp	37.5	36.1

parameter independently. Moreover, in Table 5.8 we show the accuracy for combinations of physiological parameters that yield the highest accuracy, both in 2-class and in 3-class classification.

When taking each parameter independently, the parameter that gives us the best accuracy in the two-class classification is the AccGSR with an accuracy of 70.8% for both the SVM classifier as well as the Naïve Bayes classifier. In the three-class classification, both the number of GSR peaks and the blinking rate gave the highest accuracy of 51.4 % with an SVM classifier.

For the combinations of multiple physiological parameters, we also considered the Decision trees classifier as it provides an importance metric for each feature. First, we consider the results for the two-class classification. The highest accuracy of 79.2 % was found for the combination AccGSR and the temperature variation in the forehead region with the SVM classifier. The second best accuracy of 75 % were found for two combinations: AccGSR and the number of blinks with an SVM classifier, as well as a combination of four features (i.e., blinks, blinking rate, forehead and nose temperature variations) for the decision tree classifier. As the decision tree classifier also provides the importance metric, we found the following importance for the four features: number of blinks - 0.17, blinking rate - 0.13, forehead temperature variation - 0.26, and the highest importance for the temperature variation in the nose region - 0.45.

In the case of three-class classification, we found the highest accuracy of 66.7 % with an SVM classifier when using three physiological parameters: number of blinks, blinking rate, and the temperature variation in the nose region. The second highest accuracy of 65.3 % was found, again with an SVM classifier, when using the GSR features (i.e., AccGSR and the number of peaks) and the temperature variation in the nose and forehead regions.

TABLE 5.8: Classification accuracies for combinations of physiological parameters

2-Class Classification			
Parameter	Classification Algorithm		
	SVM [%]	Naïve Bayes [%]	Decision Trees [%]
AccGSR; Blinks	75.0	66.7	56.3
AccGSR Forehead temp	79.2	66.7	62.5
Blinks Forehead temp	64.6	58.3	72.9
AccGSR; Blinks Forehead temp	68.8	72.9	62.5
Blinks; Forehead temp Nose temp	70.8	64.6	72.9
Blinks; Blink rate Forehead temp Nose temp	62.5	60.4	75
3-Class Classification			
Parameter	Classification Algorithm		
	SVM [%]	Naïve Bayes [%]	Decision Trees [%]
Peaks; Blinks Blink rate	61.1	43.1	54.2
Blinks; Blink rate Nose temp	66.7	41.7	61.1
AccGSR; Peaks Forehead temp Nose temp	65.3	50.0	43.1
Blinks; Blink rate Forehead temp Nose temp	61.1	33.3	47.2

Therefore, the most important features that differentiate the two conditions are the GSR features. For the combination of different physiological parameters, the temperature variation in the forehead region improves the overall performance. Thus, by using the AccGSR in combination with the temperature variation in the forehead region we can obtain classification accuracy of 79.2 %.

5.5 Discussion

This study was designed with three conditions in order to investigate the effects of two interaction styles of a humanoid robot and the absence of the robot on the performance and the variation of some physiological parameters for 24 participants.

The physiological parameters considered are two GSR features (i.e., AccGSR and the number of peaks), two blink features (i.e., total number of blinks and the blink rate), and the facial temperature variation in three regions (i.e., nose, forehead and perinasal). The task used in this experiment is a well known laboratory stressor (i.e., the Stroop task). This task was chosen as it requires the participants to really pay attention to the task at hand. As seen from the almost non-existent errors committed, the participants were very careful not to commit any mistakes. The first condition was always the Control condition, in which the participants had to perform the task without the robot. The following two conditions were randomized between participants and consisted in a different behavior displayed by the robot. The robot displayed either an encouraging or a stressing behavior. At every four seconds it would give a verbal content (either stressing or encouraging). While we were expecting our participants to perceive the encouraging robot as encouraging, we found that only half of the participants considered it encouraging, while for the stressing behavior, only three participants did not perceive it as stressing.

One limitation of our work is that the robot gives the feedback no matter what the participant does. We believe that is one of the reasons why the participants perceived the robot as stressing instead of encouraging. The robot should provide any verbal feedback only after the participant made a selection, and the verbal content should be in accordance with the performance of the participants.

We found that there is a slight interaction between the time of the day when the task was performed, the ME-type and the task performance. We believe this result is of interest for social robotics. A robot that knows the ME-type of the individual it interacts with, can know when is the optimum time of the day to propose certain activities. This has implications for any assistive and education applications of social robotics. One limitation concerning this result is related to the fact that we divided the interaction times into either morning (before 15h00) or evening (after 15h00). We believe a better categorization would be morning (between 7h00 and 11h00), afternoon (between 11h00 and 16h00), and evening (after 16h00). A novel study is currently designed with this division of the interaction time. Furthermore, we did not consider the Intermediate ME-type. The study of Adan (Adan, 1991) has shown that most research related to ME-type considers only morning and evening type individuals and recommends the inclusion of Intermediate type individuals in the analysis.

Our results show that morning type individuals are associated mostly with the two quadrants of the sensory profile that rely on a low neurological threshold (meaning, they need low intensity stimuli), while for the evening type individuals an association was found with the quadrants that rely on a high neurological threshold (meaning, they require high intensity stimuli). This result is of potential application in a social interaction between an individual and a robot. Furthermore, knowing the ME-type

of the individual it interacts with, beside knowing when it should propose different activities, a social robot can also know the intensity of stimuli it should use throughout the interaction. This is a result that we did not find in previous research related to social robotics. We believe that further investigation is required in order to better understand the full potential of this result.

For the Encouraging robot, the designed HRI interface was a success. According to Yanco (Yanco et al., 2004), a successful HRI interface should lower the cognitive load on the user. Our results show that during the Encouraging robot condition, the cognitive load of our participants was lower than during the Stressing robot condition.

5.6 Conclusions

In this chapter we have presented a within participants study that investigated the effects of two different interaction styles of a humanoid robot on the performance and the variation of some physiological parameters of 24 participants. Three conditions were developed for this study (Control - with no robot, Encouraging Robot, and Stressing Robot). The experimental task consisted of the non-verbal word-color Stroop Task. The physiological parameters investigated are the galvanic skin response, the blinks, and the facial temperature variation.

Our results show that the physiological parameter that can be used to distinguish between the two interaction styles of the robot is the AccGSR. The Stressing Robot led to a significantly higher values for the AccGSR than the Encouraging Robot. Furthermore, by using either a Naïve Bayes classifier or an SVM classifier, we obtained an accuracy of 70.8% for the two-class classification. When combining the AccGSR with the temperature in the forehead region and by using a Naïve Bayes classifier we obtained an accuracy of 79.2% for the two-class classification. Regarding the user profile, significant results were found for the personality trait of extraversion and for the sensory profile.

5.7 Contribution

The contribution of this chapter mainly consists in the usage of the morningness-eveningness type (ME- type) of an individual in HRI and the usage of sensory profile. We found an interaction between the ME-type and the time of the day, which can be of great importance in the field of social robotics. Knowing when to propose certain activities to the individual it interacts with can lead to an increased acceptance of the robotic system. Furthermore, our results show that there is a relationship between the ME-type and the sensory profile, which was not discussed before in the literature.

This work was published in the International Journal of Social Robotics (Agrigoroaie et al., 2019), “Cognitive Performance and Physiological Response Analysis: Analysis of the Variation of Physiological Parameters based on User’s Personality, Sensory Profile, and Morningness-Eveningness type in a Human-Robot Interaction Scenario.”

Chapter 6

RQ2: Relationship between ME-type and time of the day in relation to cognitive performance

6.1	Introduction	54
6.2	Experimental design	54
6.2.1	Experimental platform	54
6.2.2	Questionnaires	57
6.2.3	Scenario and experimental task	58
6.2.4	Participants	59
6.2.5	Interaction time	60
6.3	Results	60
6.3.1	Stroop game	61
	Task performance based on different factors	61
	Results for morning type individuals	66
	Results for evening type individuals	67
	Results for intermediate type individuals	68
	Other results	69
6.3.2	Integer Matrix	71
	Task performance based on different factors	71
	Results for morning type individuals	73
	Results for evening type individuals	74
	Results for intermediate type individuals	74
	Other results	74
6.3.3	Decimal Matrix	75
	Task performance based on different factors	75
	Results for morning type individuals	77
	Results for evening type individuals	77
	Results for intermediate type individuals	78
	Other results	78
6.4	Discussion and the message to take home	78
6.5	Conclusion	79
6.6	My contribution	79

6.1 Introduction

Based on the results that we obtained in Chapter 5, we have decided to further investigate the relationship that exists between ME-type and the time of the day when a task is performed and its influence on cognitive performance. For this purpose, we developed an online platform where participants could create an account and perform three cognitive tasks, at their own pace. Participants were recruited through Facebook, SurveyTandem, SurveyCircle, and SurveySwap, as well as by sending emails to colleagues, friends, and family.

In this chapter, we present the results of the 135 participants who filled out all the questionnaires on our online platform.

Therefore, the main research question for this study is **RQ1: Is there a relationship between morningness-eveningness and the time of the day when a cognitive task is performed?** While the study presented in Chapter 5 shows that such a relationship does exist, the participants performed the task when instructed by the experimenter, and not whenever they preferred. Therefore, in this study, we investigated this relationship by taking into consideration that the experimental task was performed whenever the participants preferred (i.e., at what hour of the day they preferred).

6.2 Experimental design

First, we start by presenting the online platform that we created. We then continue with the presentation of the experimental tasks and the description of the participants who performed the tasks.

6.2.1 Experimental platform

For this study, a custom web-based application was developed using CakePHP¹ and connected to an MySQL² database. The study is available at: <http://experiment.dotpxl.ro>.

The platform was designed to be modular, so as to be easily extended for future studies as well. Each study (or experiment) is made up of multiple questionnaires and of one or multiple experimental tasks.

When creating an account, each participant has to provide all of the following details (see also Figure 6.1):

- Username
- Age group (0-18, 19-25, 26-35, 36-50, 51-65, and over 65 years old)
- Gender (male or female)
- Country (country of residence or birth)
- Studies
- Email
- Password

¹<https://book.cakephp.org/2.0/en/index.html>

²<https://www.mysql.com/>

FIGURE 6.1: Registration page

Besides these elements, we showed each participant three images (also shown in Figure 6.1), and they had to select if the images indicated a positive, negative, or neutral feeling. The purpose of this was to validate the three images for future studies.

Once the account was created, a confirmation email was sent to the email address provided by the participants. Each participant has access to the following pages:

- **Dashboard:** on this page, the participants can see the results from the questionnaires, the variation of the performance from the experimental tasks as well as the leader board for each game (see Figure 6.2). In the leader board, we display the top three scores among all participants, as well as the position and the best score of the logged it participant.
- **Experiments page:** on this page, the participants have access to a brief description of the experiment, the questionnaires that need to be completed, and the experimental tasks (see Figure 6.3). Furthermore, on the top of the page, we display a message informing the participant either what tasks still have to be performed in order to complete the study, or a message informing the participant that they successfully completed the study.
- **Study statistics page:** on this page different statistics regarding the participants are displayed (e.g., total number of participants with accounts, age distribution, where the participants are from, etc.) (see Figure 6.4).
- **Contact:** a page where the participants can send a message to the experimenter.
- **About:** a page where the participants can find out more about our work.

In order to ensure ourselves that each participant filled all study questionnaires, the experimental tasks are disabled until all questionnaires are completed.

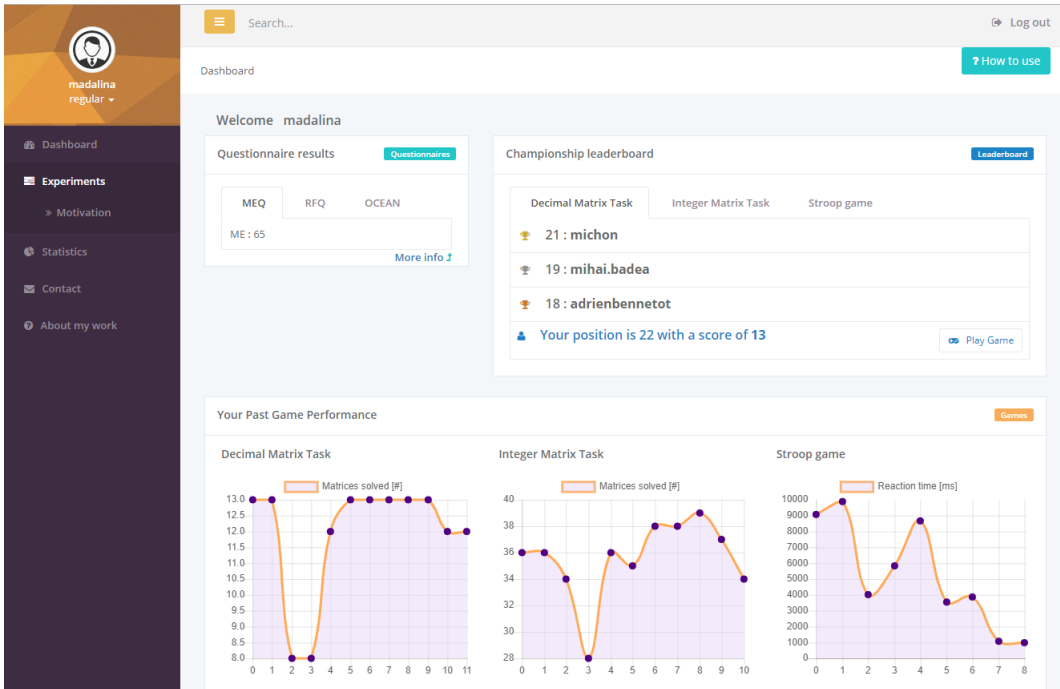


FIGURE 6.2: User dashboard

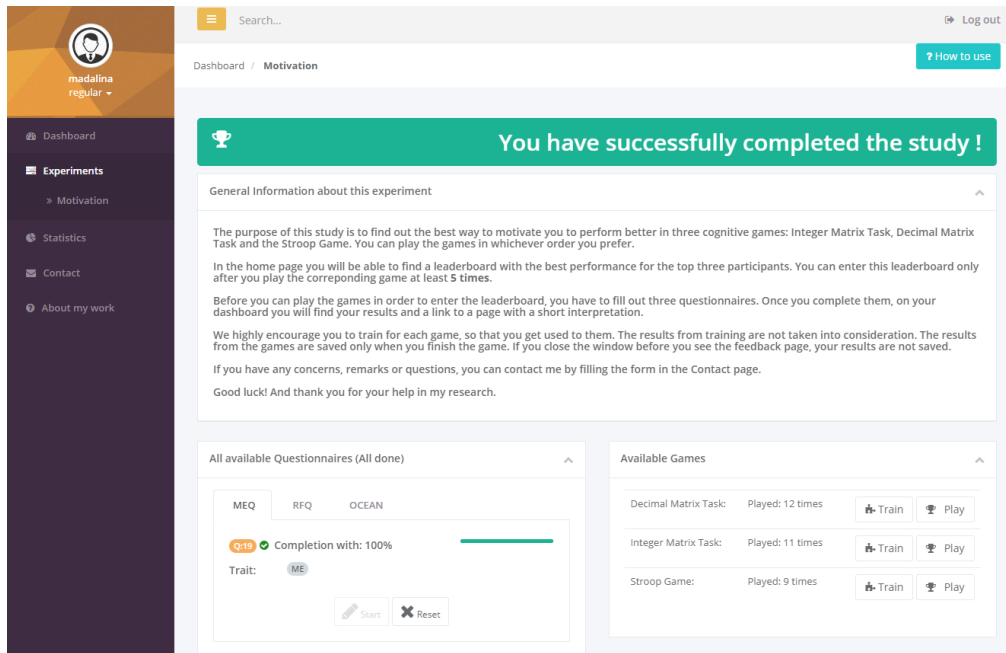


FIGURE 6.3: Experiments page

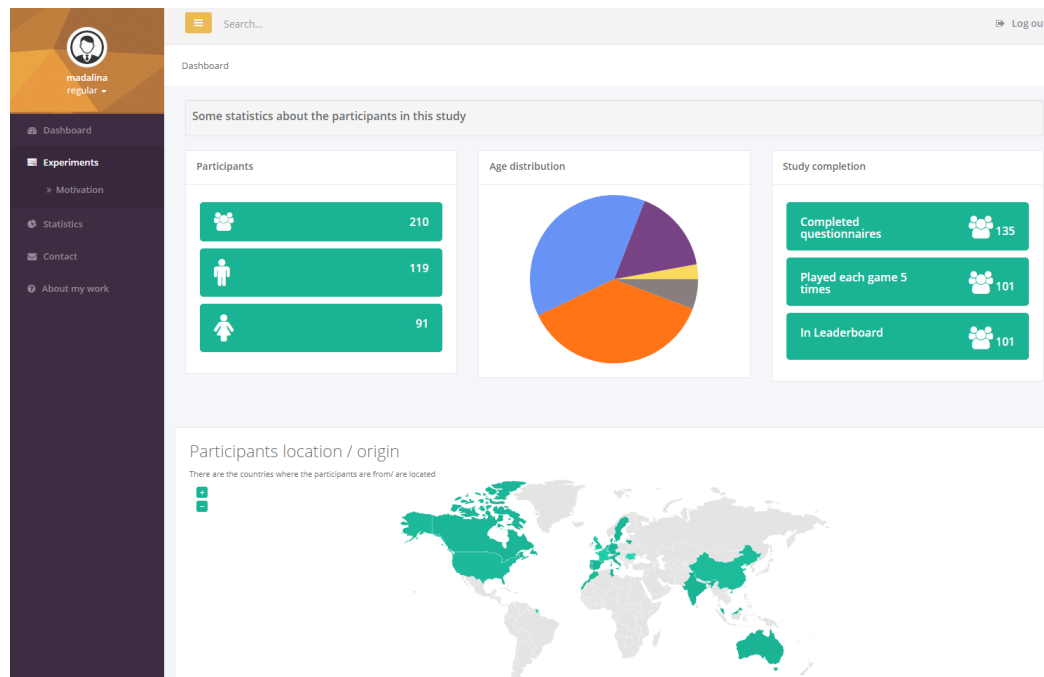


FIGURE 6.4: Study statistics page

6.2.2 Questionnaires

Each participant had to fill out the Morningness-Eveningness Questionnaire (MEQ) (presented in Chapter 4.5), the BIG5 personality questionnaire, and the Regulatory Focus Questionnaire - Proverb Form.

Based on the scores from MEQ, the participants can be divided into 5 categories (i.e., definite evening, moderate evening, intermediate, moderate morning, definite morning). Most studies (see (Adan et al., 2012) for a review) consider only two categories (ME-type), either morning type or evening type. Therefore, based on the MEQ score, we divided our participants into morning type (ME-M, scores greater than 49) and evening type (ME-E, scores less than or equal to 49). However, according to (Adan, 1991), it is important to consider for analysis also the Intermediate type individuals. As a result, we have decided to also split the participants into three groups (MIE-type): morning type (MIE-M, scores greater than 58), intermediate type (MIE-I, scores greater than 41 and less than or equal to 58), and evening type (MIE-E, scores less than or equal to 41). The distribution of the participants based on the ME-type and MIE-type is shown in Table 6.1. As can be seen, the majority of the participants fall into the intermediate category (i.e., 69 out of 122 participants).

BIG5 personality questionnaire (BIG5)(Goldberg, 1990) is one of the most used personality questionnaires in HRI. It measures five personality traits: Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism. The questionnaire is also known as the OCEAN personality questionnaire. The questionnaire is based on the factor analysis of responses to personality items. The test consists of 48 questions each with a five-point Likert scale (1 - “Strongly disagree”, 5 - “Strongly agree”). According to (Barrick et al., 1991), traits that are associated to Openness (BIG5-O) include being imaginative, curious, original, artistically sensitive. Conscientiousness (BIG5-C) is associated with being careful, thorough, responsible, organized or hardworking. Traits associated with Extroversion (BIG5-E) include being sociable, assertive or talkative. Agreeableness (BIG5-A) is associated

with being courteous, flexible, trusting, cooperative, while Neuroticism (BIG5-N) is usually associated with traits like begin emotional, worried, insecure. For each of the five personality traits scores range between 1 and 5. We have divided participants into two groups (i.e., low level and high level) for each trait based on the average score found in our population. The distribution of the participants based on the results of this questionnaire is shown in Table 6.1.

Regulatory Focus Questionnaire - Proverb Form (Faur et al., 2017) (RFQ) was developed in French and it contains 18 items (proverbs). The authors also provide in (Faur et al., 2017), the English equivalent for each of the 18 proverbs. The questionnaire measures two traits: Promotion and Prevention. Scores for each trait range between 1 and 7. The predominant trait is considered the one with the highest score. According to Higgins (the initiator of the theory), the trait can be either Promotion or Prevention, or strengths of the predominant trait can be taken into account. While this questionnaire was given to the participants, the analysis based on the regulatory focus predominant trait is outside the scope of this study.

6.2.3 Scenario and experimental task

For this study, each participant had to first fill out the three questionnaires (MEQ, BIG5, and RFQ) available on the Experiments page. As previously mentioned, the experimental tasks were disable if the three questionnaires were not completed. However, each participant could train for each of the experimental tasks. In the training session, the duration of the games were half of that during the tests, and while the performance was recorded in the dataset, it did not influence their overall score and it was not shown in the Past Performance section of the Dashboard page.

The experimental tasks consisted in three cognitive games: Stroop game, Integer Matrix and Decimal Matrix. The three games were already presented in Chapter 3.5.

The Stroop game consisted of 40 trials: half congruent and half incongruent, presented in a random order. The performance measures for this game are: the reaction time during the congruent trials (RTCong), the reaction time during the incongruent trials (RTIncong), the reaction time during all trials (RTTotal), the number of correct trials, the number of wrong trials, and the number of trials for which no answer was given. For each trial, the participants had 3 seconds to answer, after which a new trial was shown. The order of the button was randomized between the trials. We have also defined a performance measure for the leader board which is defined as: the average time needed for the correct trials, to which we added the average time needed for the wrong trials, and if there were trials to which no answer was given, we added 3 seconds, which is the time the participants had to provide an answer.

The Integer and Decimal Matrix Tasks had a total duration of 90 seconds. The difficulty level for the two games was set to medium (i.e., a matrix size of 3x3). Decimal Matrix Task featured digits with only one decimal. The performance measures for these two games are: the time needed to solve a matrix, the number of correct matrices, and the number of mistakes made. The leader board performance measure is represented by the number of correctly solved matrices.

The participants could play the games at any time of the day they preferred, on any device (i.e., phone, tablet or computer) and during how many days they wanted. They were instructed to play each of the three games at least five times. Considering that the BIG5-C personality trait is associated with being thorough and responsible, one of our research questions is **RQ2: Does the BIG5-C personality trait influence how many times each game was played?**. We expect that

TABLE 6.1: Distribution of the participants

Age group				
0 - 18	19 - 25	26 - 35	36 - 50	51 - 65
4	36	55	22	5
Gender		Country		
Male	Female	Romania	France	Other
61	61	62	22	38
BIG 5 Results				
Agreeableness	Neuroticism	Extroversion	Openness	Conscientiousness
Min: 2.3	Min: 1.0	Min: 1.7	Min: 2.2	Min: 2.2
Max: 4.7	Max: 4.8	Max: 4.7	Max: 4.8	Max: 5.0
Mean: 3.54	Mean: 2.87	Mean: 3.06	Mean: 3.64	Mean: 3.52
Low: 60	Low: 62	Low: 64	Low: 66	Low: 57
High: 62	High: 60	High: 58	High: 56	High: 65
MEQ results				
ME	ME	MIE	MIE	MIE
Morning type	Evening type	Morning type	Intermediate	Evening type
59	63	29	69	24

participants with high levels of BIG5-C will play more games than the participants with low BIG5-C levels.

6.2.4 Participants

A total of 210 participants (91 females) created accounts on our online platform. Out of them, 135 filled out all the questionnaires, 101 played each of the three games at least 5 times, and 109 participants played at least one game five times, 128 participants played at least one game one time, while 122 participants played at least one game three times. For the analysis, for our first research question, we are going to use the data from the 122 participants who played at least one game three times, while for our second research question we are going to use the data from all 135 participants who filled out all the questionnaires.

The distribution of the 122 participants based on their demographic information as well as the results from the questionnaires is shown in Table 6.1.

The majority of the participants are from Romania and France (62 and 22, respectively). The participants are also from: the United Kingdom (UK) (N=10), the Netherlands (N=7), and China (N=3). There are two participants from each of the following countries: Belgium, Canada, the United States of America, and Italy. Furthermore, there is one participant from each of the following countries: Australia, Austria, Denmark, Germany, India, Lithuania, Morocco, Spain, Sweden, and Tunisia.

One of the details that the participants provided when creating the accounts is represented by their studies. Their responses show that the studies our participants have completed are very variate: some are still students (high school, Bachelor's, Master's, or PhD)(N=60), some completed engineering schools (N=18), other, computer science programs (N=6), as well as psychology, communication, medicine, journalism, marketing and geography, among others.

TABLE 6.2: Distribution of the played games at different times of the day

Game	Morning	Afternoon	Evening	Night
Stroop	154	167	126	124
Integer Matrix	172	165	135	137
Decimal Matrix	148	176	123	158

6.2.5 Interaction time

The 122 participants played a total of 1785 games (Stroop - 571; Integer Matrix task - 609; Decimal Matrix task - 605). The games could be played at whatever time of the day the participants wanted. We divided the interaction times into four categories depending on the hour when the interaction took place (we took into account the participants' local hour when the results were saved in the database). The four categories are as follows:

- **morning:** interaction time between 06h00 and 12h59
- **afternoon:** interaction time between 13h00 and 16h59
- **evening:** interaction time between 17h00 and 20h59
- **night:** interaction time between 21h00 and 05h59

The distribution of the played games into the four interaction time categories are shown in Table 6.2.

6.3 Results

Taking into consideration the results from Chapter 5, and our research questions, the analysis was performed for each game separately and it was divided into multiple parts:

- Task performance based on different factors (i.e., time of the day, ME-type, MIE-type)
- Results for morning type individuals (ME-M type, MIE-M type)
- Results for evening type individuals (ME-E type, MIE-E type)
- Results for intermediate type individuals (MIE-I type)
- Influence of BIG 5-C on number of games played

The task performance is measured as: for the Stroop game (RTCong - reaction time during congruent trials, RTIncong - reaction time during incongruent trials, RTTotal - reaction time during all trials, correct answers, wrong answers, no answers), while for Integer and Decimal Matrix (time to solve a matrix, correct matrices, and wrong matrices).

We begin our analysis with the Stroop game (as this task was also investigated in Chapter 5), and then continue with Integer Matrix, and finally with Decimal Matrix.

The games were played at least three times by each participant, however, they were not played by each participant at distinct times of the day. Therefore, we cannot

TABLE 6.3: Distribution of the played Stroop games at different times of the day based on ME-type and MIE-type

Type	Stroop Game			
	Morning	Afternoon	Evening	Night
ME-type Morning	29	37	18	18
ME-type Evening	27	24	27	26
MIE-type Morning	15	10	13	10
MIE-type Intermediate	28	38	24	24
MIE-type Evening	13	13	8	10

apply a repeated measures type of analysis. The majority of our participants played all games at the same time of the day (around 80 participants played all three games in one session). Thus, we have decided to analyse only the two best performances of each participant. We should point out again, that the participants freely chose the time when to play these games.

We have selected the two performances for each participant for which they had the highest leader board performance measure. More specifically, for the Stroop game, we define the performance measure as the sum of the average reaction time for correct answers, the average reaction time for wrong answers and if there were trials for which the participant did not give an answer, we added 3 seconds, which is the time they had to give an answer. For the Integer and Decimal Matrix games, the performance measure that we considered is the number of correctly solved matrices.

For the analysis, we first applied a Shapiro-Wilk normality test in order to determine if parametric or non-parametric tests can be applied. If the data is normally distributed, we applied an ANOVA-type analysis with pairwise t-tests to determine differences between groups. For the non-normally distributed data, we applied a Kruskal-Wallis test, with pairwise Wilcoxon test to determine differences between the groups. For all pairwise tests (i.e., t-test and Wilcoxon test) we applied a Bonferroni correction.

6.3.1 Stroop game

A total of 571 games were played by 109 participants. Out of these, we only selected the participants who played the game at least three times (i.e., in order to eliminate the influence of novelty from the results). Therefore, at the end, for the analysis we used 206 games played by 103 participants. The games were played at all hours of the day and by all types of individuals based on the results from MEQ (see also Table 6.3).

Task performance based on different factors

The significant results based on the three investigated factors (i.e., time of the day, ME-type and MIE-type) are shown in Table 6.4.

The time of the day influences the performance during the incongruent trials (RTIncong), with a better performance during the morning hours compared to all other hours of the day (see Figure 6.5). A pairwise t-test comparisons with Bonferroni correction shows that during the morning hours individuals are faster than during the afternoon hours ($p = 0.042$, $M_{morning} = 1.26$, $M_{afternoon} = 1.33$), are faster than

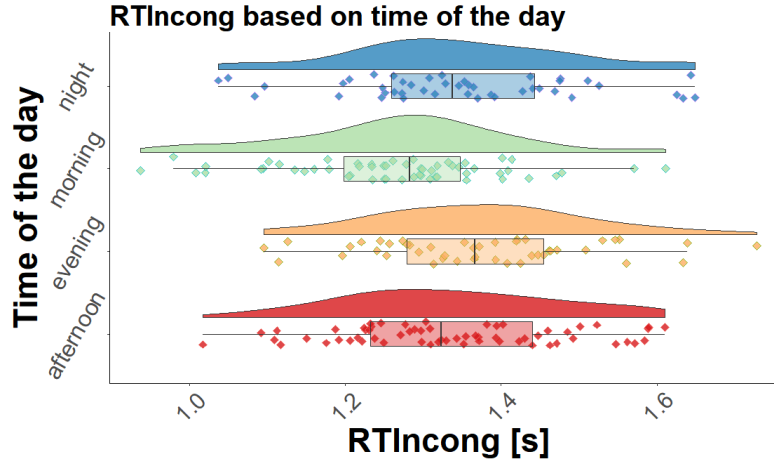


FIGURE 6.5: Reaction time during incongruent trials based on the time of the day for Stroop Game

TABLE 6.4: Significant results for task performance based on different factors for Stroop Game

Dependent variable	Factor	F value / χ^2	p-value
RTIncong	time of the day	$F(3, 202) = 5.48$	0.001 **
Correct	time of the day	$\chi^2 = 11.02$	0.01 *
Wrong	time of the day	$\chi^2 = 9.79$	0.02 *
RTCong	MIE-type	$\chi^2 = 18.11$	0.0001 ***
RTIncong	MIE-type	$F(2, 203) = 7.72$	0.0005 ***
RTTotal	MIE-type	$\chi^2 = 15.21$	0.0004 ***
NoAnswer	MIE-type	$\chi^2 = 6.59$	0.036 *

during the evening hours ($p = 0.001$, $M_{evening} = 1.37$), and faster than during the night hours ($p = 0.03$, $M_{night} = 1.34$). The same tendency was found for the reaction time during the congruent trials ($RT_{Cong} - M_{morning} = 1.20$, $M_{afternoon} = 1.23$, $M_{evening} = 1.25$, $M_{night} = 1.24$), and for all trials as well ($RT_{Total} - M_{morning} = 1.23$, $M_{afternoon} = 1.28$, $M_{evening} = 1.31$, $M_{night} = 1.29$). However, the results for RT_{Cong} and RT_{Total} are not statistically significant.

No significant results were found for the ME-type of our participants. This confirms that it is important to take into consideration the MIE-I type individuals as well.

When considering the MIE-type of the individuals, evening type individuals are faster than morning type individuals ($RT_{Cong} - p = 0.0008$, $M_{MIE-E} = 1.19$, $M_{MIE-M} = 1.30$; $RT_{Incong} - p = 0.0009$, $M_{MIE-E} = 1.28$, $M_{MIE-M} = 1.39$; $RT_{Total} - p = 0.0014$, $M_{MIE-E} = 1.24$, $M_{MIE-M} = 1.34$) and they have fewer trials for which they gave no answer ($p = 0.047$, $M_{MIE-E} = 1.5$, $M_{MIE-M} = 3.52$). However, when investigating the wrong answers, morning type individuals have a tendency to have less wrong answers ($M_{MIE-M} = 0.47$, $M_{MIE-E} = 0.68$).

Intermediate type individuals are faster than morning type individuals ($RT_{Cong} - p = 0.0002$, $M_{MIE-I} = 1.21$; $RT_{Incong} - p = 0.0039$, $M_{MIE-I} = 1.31$; $RT_{Total} - p = 0.0021$, $M_{MIE-I} = 1.34$). Morning type individuals are slower, but they also make fewer mistakes. Evening and intermediate type individuals have very similar

performance measures.

Next, we investigated if there is an interaction between the time of the day when the task was performed and the MIE-type of the individuals.

In Figure 6.6 and Figure 6.7 are shown the interaction plots for each of the six performance measures (RTCong, RTIncong, RTTotal, correct trials, wrong trials, and trials with no answer).

For RTCong (Figure 6.6a), morning-type individuals, have lower reaction times than both evening-type individuals and intermediate-type individuals. There is no interaction between the groups (morning vs evening and morning vs intermediate). However, intermediate type individuals are faster only in the morning hours than evening type individuals (in the morning - $M_{MIE_I} = 1.15$, $M_{MIE-E} = 1.21$), while at all other times of the day, evening type individuals are faster than intermediate type individuals (in the afternoon - $M_{MIE-I} = 1.23$, $M_{MIE-E} = 1.20$; in the evening - $M_{MIE-I} = 1.21$, $M_{MIE-E} = 1.18$; at night - $M_{MIE-I} = 1.25$, $M_{MIE-E} = 1.16$).

For the incongruent trials (Figure 6.6b), again, morning-type individuals have lower reaction times than both evening and intermediate type individuals (with no interaction between the groups). As with the congruent trials, during the morning hours, intermediate type individuals are faster than evening type individuals (in the morning - $M_{MIE-I} = 1.24$, $M_{MIE-E} = 1.27$), while at the other times of the day, evening type individuals are faster than intermediate type individuals (in the afternoon - $M_{MIE-I} = 1.33$, $M_{MIE-E} = 1.30$; in the evening - $M_{MIE-I} = 1.33$, $M_{MIE-E} = 1.31$; at night - $M_{MIE-I} = 1.35$, $M_{MIE-E} = 1.25$).

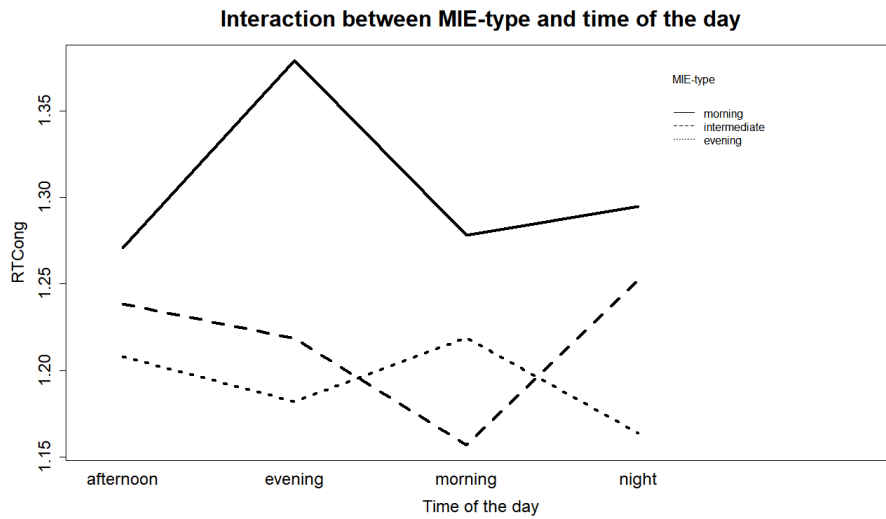
For all trials (Figure 6.6c), the same result was found as with the congruent and incongruent trials. In the morning - $M_{MIE-I} = 1.19$, $M_{MIE-E} = 1.24$; in the afternoon - $M_{MIE-I} = 1.28$, $M_{MIE-E} = 1.25$; in the evening - $M_{MIE-I} = 1.27$, $M_{MIE-E} = 1.24$; at night - $M_{MIE-I} = 1.30$, $M_{MIE-E} = 1.20$.

For the number of correct trials (Figure 6.7a), the number of correct answers is between 36 and 40, for all times of the day and all MIE-types, except for the morning type individuals who played the games in the evening, when the number of correct answers is $M_{MIE-M} = 31$.

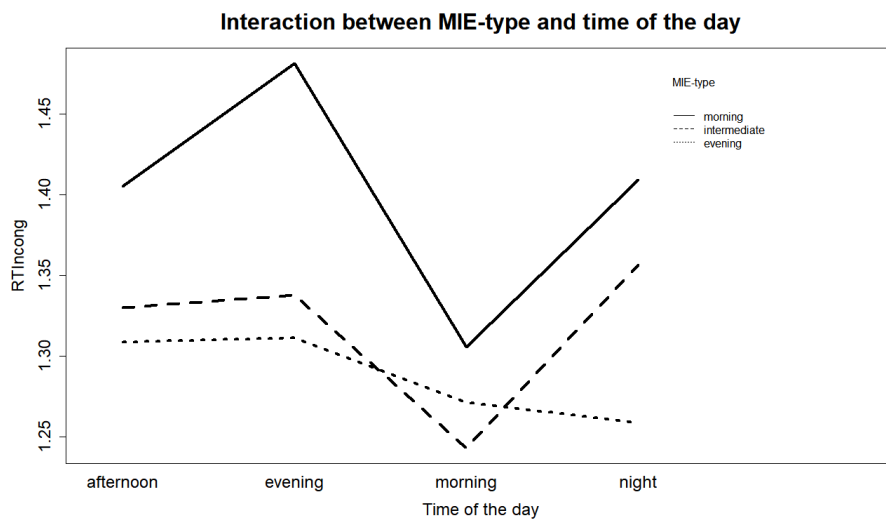
The number of wrong trials (Figure 6.7b), is mostly below 1, with the best performance for morning type individuals in the afternoon ($M_{MIE-M} = 0.10$). The performance in the morning, at night and in the evening is comparable for all MIE-type individuals. More specifically, at night, in the evening and in the morning, evening type individuals made fewer mistakes than intermediate type individuals. However, in the afternoon, evening type individuals have more mistakes ($M_{MIE-E} = 1.07$) than intermediate type individuals ($M_{MIE-I} = 0.73$). Morning type individuals made fewer mistakes than evening and intermediate type individuals in the afternoon, and at night, while in the morning they made more mistakes than the other two types of individuals.

Considering the last performance parameters, number of trials without answer, (Figure 6.7c), the performance for evening and intermediate type individuals is very similar, during all hours of the day. For the morning type individuals, the time of the day with the worst performance is represented by the evening ($M_{morning} = 7.76$), followed by the night hours ($M_{night} = 2.8$).

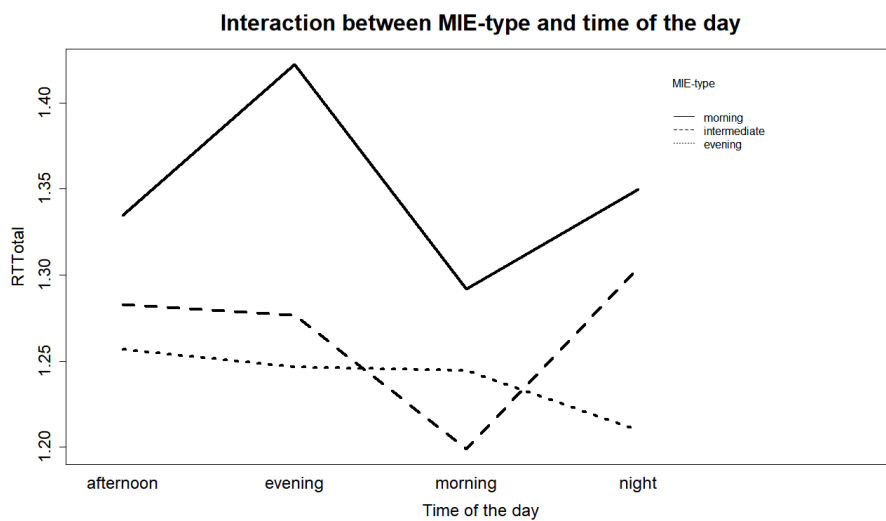
Discussion: Taking all these results into consideration and the interaction plots in Figure 6.6 and Figure 6.7, morning type individuals have the worst performance in the evening hours, followed by the night and afternoon hours. Except for the number of wrong answers, morning type individuals have the best performance in the morning hours. Intermediate type individuals are only faster than evening type individuals in the morning, have more correct answers than evening type individuals



(A) Congruent trials

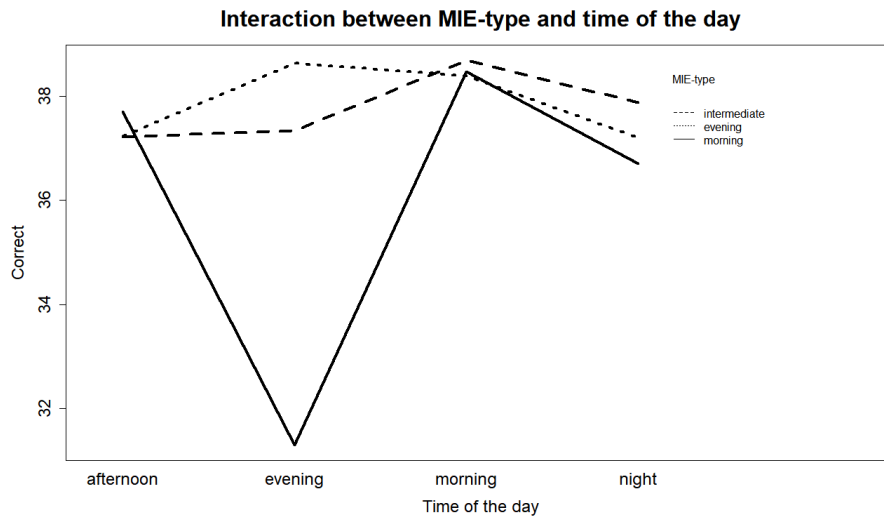


(B) Incongruent trials

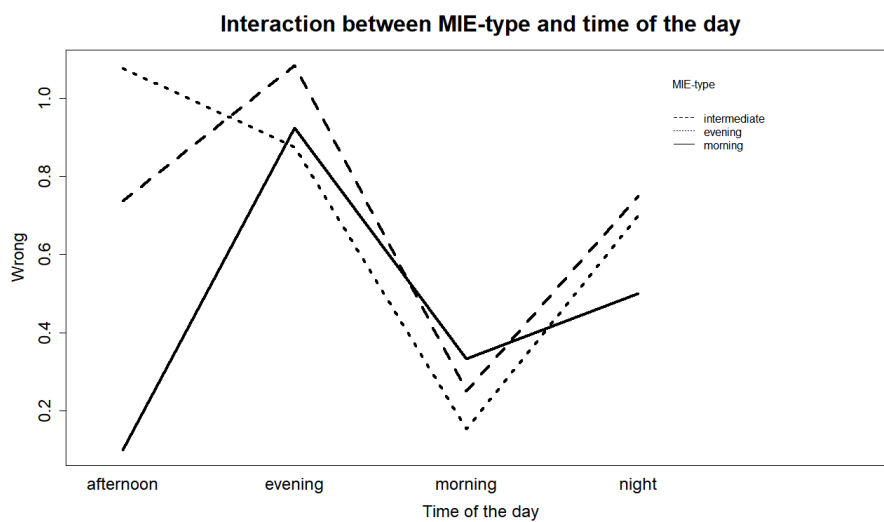


(C) All trials

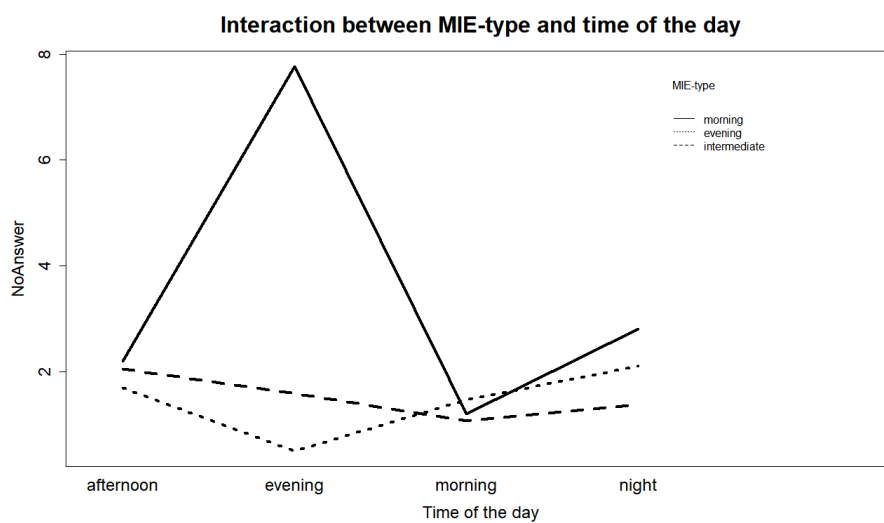
FIGURE 6.6: Interaction plot for MIE-type and the interaction time for Stroop Game



(A) Correct



(B) Wrong



(C) No Answer

FIGURE 6.7: Interaction plot for MIE-type and the interaction time for Stroop Game

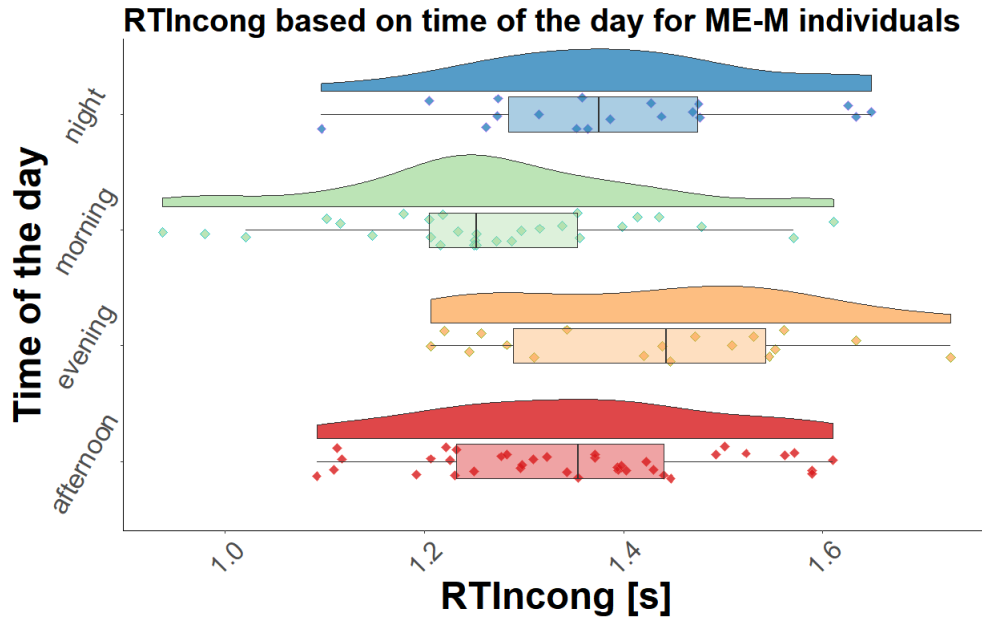


FIGURE 6.8: Reaction time for incongruent trials for ME-M type individuals based on the time of the day for Stroop Game

in the morning and at night, have less wrong answers than evening type individuals only in the afternoon, and have less trials with no answers in the morning and at night.

Results for morning type individuals

First, we investigate the results of the ME-M type individuals (i.e., when dividing individuals into either morning or evening types). Significant differences depending on the time of the day were found for RTIncong ($F(3, 98) = 5.16, p = 0.0023$), RTTotal ($F(3, 98) = 3.71, p = 0.01$), and the number of correct answers ($\chi^2 = 9.10, p = 0.027$).

For the incongruent trials (RTIncong), in the morning, individuals are significantly faster than in the evening ($p = 0.003, M_{morning} = 1.26, M_{evening} = 1.42$). Also, in the morning hours individuals are faster than during the night hours ($p = 0.033, M_{night} = 1.39$). This result can also be seen presented in Figure 6.8.

For all trials, individuals are significantly faster in the morning only compared to the evening hours ($p = 0.015, M_{morning} = 1.23, M_{evening} = 1.37$). Moreover, individuals are faster in the morning than at any other time of the day.

Lastly, another significant result was found for the number of correct answers. Morning type individuals have the least number of correct answers in the evening ($M_{morning} = 38.51; M_{afternoon} = 37.32; M_{evening} = 33.33; M_{night} = 37.05$). However, the differences between the groups are not significant.

Next, we have selected only the morning type individuals, based on the instructions from the Morningness-Eveningness Questionnaire (i.e., MEQ scores greater than 58). There are 24 participants of MIE-M type. A significant result was found only for the reaction time during the incongruent trials based on the time of the day (see Figure 6.9), $F(3, 44) = 3.15, p = 0.03$. A pairwise t-test with Bonferroni correction revealed significant differences between the morning ($M_{morning} = 1.30$) and the evening ($M_{evening} = 1.48$) hours. As can be seen in Figure 6.9, in the morning, MIE-M individuals are faster than during any other time of the day.

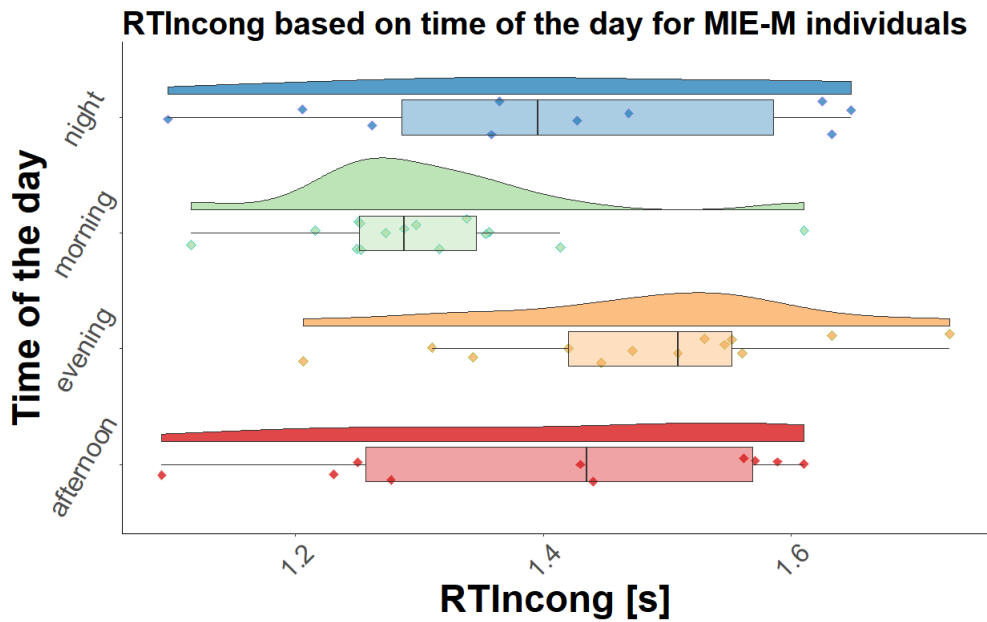


FIGURE 6.9: Reaction time for incongruent trials for MIE-M type individuals based on the time of the day for Stroop Game

TABLE 6.5: Results for ME-E individuals for Stroop Game

Measure	Morning time	Afternoon time	Evening time	Night time
RTCong	$M = 1.202$	$M = 1.233$	$M = 1.214$	$M = 1.213$
RTIncong	$M = 1.267$	$M = 1.317$	$M = 1.338$	$M = 1.313$
RTTotal	$M = 1.234$	$M = 1.274$	$M = 1.275$	$M = 1.263$
Correct	$M = 38.59$	$M = 37.25$	$M = 37.48$	$M = 37.73$
Wrong	$M = 0.259$	$M = 1.041$	$M = 1.074$	$M = 0.730$
NoAnswer	$M = 1.148$	$M = 1.708$	$M = 1.444$	$M = 1.538$

Results for evening type individuals

Next, we investigate the results for the evening type individuals. First, we analyse the results for the ME-E type individuals. However, no significant results were found for the performance based on the time of the day. In Table 6.5, we show the results of the ME-E type individuals. As can be seen, evening type individuals have the best performance during the morning hours, followed by the the night hours and then evening hours. The worst performance was obtained during the afternoon hours. It is important to point out that these are just tendencies, as none of these results are significant.

No significant results were found for MIE-E type individuals either. Based on the results from Table 6.6, MIE-E type individuals are the fastest at night, followed by the evening hours, however, they have the most correct answers during the evening hours followed by the morning hours, the least number of mistakes in the morning followed by the night and evening hours, and the least number of trials with no answer during the evening hours. Taking all these into consideration, we can say that MIE-E type individuals have in general, better results in the evening hours, followed by the morning hours. The worst performance was found for the afternoon hours.

TABLE 6.6: Results for MIE-E individuals for Stroop Game

Measure	Morning time	Afternoon time	Evening time	Night time
RTCong	$M = 1.218$	$M = 1.207$	$M = 1.182$	$M = 1.163$
RTIncong	$M = 1.271$	$M = 1.308$	$M = 1.311$	$M = 1.258$
RTTotal	$M = 1.244$	$M = 1.256$	$M = 1.246$	$M = 1.209$
Correct	$M = 38.38$	$M = 37.23$	$M = 38.62$	$M = 37.20$
Wrong	$M = 0.153$	$M = 1.076$	$M = 0.875$	$M = 0.700$
NoAnswer	$M = 1.461$	$M = 1.692$	$M = 0.500$	$M = 2.100$

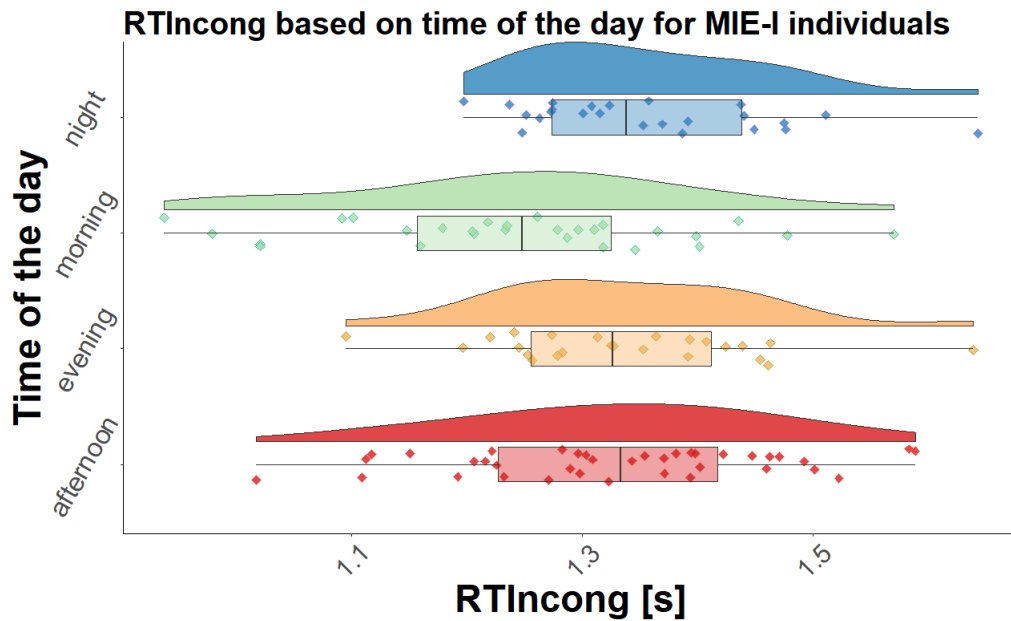


FIGURE 6.10: Reaction time for incongruent trials for MIE-I type individuals based on the time of the day for Stroop Game

Results for intermediate type individuals

Lastly, we investigated the results for the intermediate type individuals. Significant results were found for RTIncong ($F(3, 110) = 4.06, p = 0.008$), RTTotal ($F(3, 110) = 3.91, p = 0.01$) and correct answers ($\chi^2 = 8.21, p = 0.04$)

For the incongruent trials (see Figure 6.10), intermediate type individuals have a better performance in the morning than during the night ($p = 0.013, M_{morning} = 1.24, M_{night} = 1.35$), in the morning than in the afternoon ($p = 0.052, M_{afternoon} = 1.33$), and in the morning than in the evening ($p = 0.063, M_{evening} = 1.33$). No differences can be seen between the afternoon, evening and at night.

For all trials, in the morning, individuals are faster than at night ($p = 0.014, M_{morning} = 1.19, M_{night} = 1.30$), faster in the morning than in the afternoon ($p = 0.038, M_{afternoon} = 1.28$), and faster than in the evening ($p = 0.13, M_{evening} = 1.27$).

For the number of correct answers, during the morning, intermediate type individuals solve correctly more matrices than during the afternoon ($p = 0.04, M_{morning} = 38.67, M_{afternoon} = 37.21$). During the evening, and during the night, intermediate type individuals answer correctly to an average of 37 trials.

TABLE 6.7: Results from pairwise comparisons based on age group for Stroop Game

Performance Measure	Group1	Group2	p-value
RTCong	0-18 ($M = 1.20$)	51-65 ($M = 1.47$)	0.02
	19-25 ($M = 1.17$)	36-50 ($M = 1.33$)	< 0.0001
		51-65 ($M = 1.47$)	0.0004
	26-35 ($M = 1.22$)	36-50 ($M = 1.33$)	0.0004
		51-65 ($M = 1.47$)	0.0008
RTTotal	0-18 ($M = 1.29$)	51-65 ($M = 1.51$)	0.006
	19-25 ($M = 1.22$)	36-50 ($M = 1.36$)	< 0.0001
		51-65 ($M = 1.51$)	0.0002
	26-35 ($M = 1.26$)	36-50 ($M = 1.36$)	0.003
		51-65 ($M = 1.51$)	0.0002
Correct	19-25 ($M = 37.95$)	51-65 ($M = 28.12$)	0.011
	26-35 ($M = 37.74$)		0.010
Wrong	0-18 ($M = 1.66$)	36-50 ($M = 0.26$)	0.014
NoAnswer	19-25 ($M = 1.43$)	51-65 ($M = 10.75$)	0.0043
	26-35 ($M = 1.58$)		0.0040

Other results

Other significant results that were found are shown in Table 6.8. Age has an influence on all performance parameters except RTIncong, BIG5-A influences RTCong, RTTotal, correct trials and the number of trials without an answer, sleep time influences the number of correct answers and the number of trials without an answer, while how the participants felt during that day influenced all performance parameters except for the number of wrong answers.

Individuals with low levels for BIG5-A personality trait are faster during the congruent trials ($p = 0.01$, $M_{low} = 1.20$, $M_{high} = 1.25$), for all trials ($p = 0.04$, $M_{low} = 1.25$, $M_{high} = 1.30$), have more correct answers ($p = 0.0039$, $M_{low} = 37.81$, $M_{high} = 36.97$) and have fewer trials for which they gave no answer ($p = 0.01$, $M_{low} = 1.54$, $M_{high} = 2.38$) compared to individuals with high level for BIG5-A.

For the age factor, the distribution between the different age groups is not well balanced (0-18 - N=3; 19-25 - N=32; 36-50 - N=49; 36-50 - N=15; and 51-65 - N=4). In Table 6.7, we show the differences between the different age groups. However, these should be considered with caution, as the distribution is not balanced.

Considering our second research question **RQ2: Does the BIG5-C personality trait influence how many times each game was played?**, we have found that, individuals with high levels for BIG5-C played on average more games than individuals with low levels for BIG5-C ($M_{high} = 4.39$, $M_{low} = 4.08$, $\chi^2 = 0.27$, $p = 0.6$). This result is also shown graphically in Figure 6.11. As, the agreeableness personality trait proved to be significant in our previous test, we wanted to see if it does also influence the number of games played by the participants. We found that individuals with high BIG5-A levels played more games than individuals with low BIG5-A levels ($M_{high} = 4.52$, $M_{low} = 3.9$, $\chi^2 = 3.14$, $p = 0.07$).

TABLE 6.8: Significant results for task performance based on different factors for Stroop Game

Dependent variable	Factor	F value / χ^2	p-value
RTCong	age	$\chi^2 = 40.85$	< 0.0001 ***
RTTotal	age	$\chi^2 = 37.87$	< 0.0001 ***
Correct	age	$\chi^2 = 12.80$	0.012 *
Wrong	age	$\chi^2 = 10.67$	0.03 *
NoAnswer	age	$\chi^2 = 14.62$	0.005 **
RTCong	BIG5-A	$\chi^2 = 6.46$	0.01 *
RTTotal	BIG5-A	$\chi^2 = 4.13$	0.041 *
Correct	BIG5-A	$\chi^2 = 8.29$	0.003 **
NoAnswer	BIG5-A	$\chi^2 = 6.47$	0.01 *

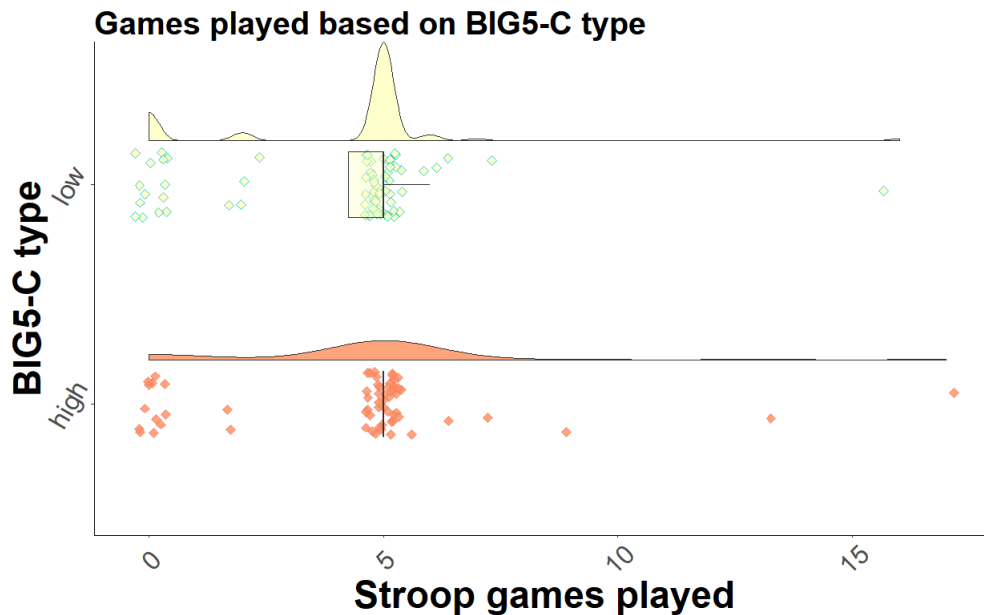


FIGURE 6.11: Stroop game played based on BIG5-C personality trait

TABLE 6.9: Distribution of the played Integer Matrix games at different times of the day based on ME-type and MIE-type

Integer Matrix Game				
Type	Morning	Afternoon	Evening	Night
ME-type Morning	34	48	23	23
ME-type Evening	41	19	39	35
MIE-type Morning	15	16	19	13
MIE-type Intermediate	42	42	27	32
MIE-type Evening	18	9	16	13

6.3.2 Integer Matrix

A total of 609 games were played by 113 participants. Out of these, we only selected the participants who played the game at least three times (i.e., in order to eliminate the influence of novelty from the results). Therefore, at the end, for the analysis, we use 262 games played by 111 participants. The games were played at all hours of the day and by all types of individuals based on the results from MEQ and BIG5 (see also Table 6.9).

Task performance based on different factors

As for the Stroop game, for the Integer Matrix, we first checked to see if the time of the day when the game was played had any influence on the three performance parameters (i.e., time to solve matrix, number of correct matrices, and number of mistakes made).

We found that time of the day influences the reaction time ($\chi^2 = 8.07$, $p = 0.04$) and the number of correct answers ($\chi^2 = 9.66$, $p = 0.02$).

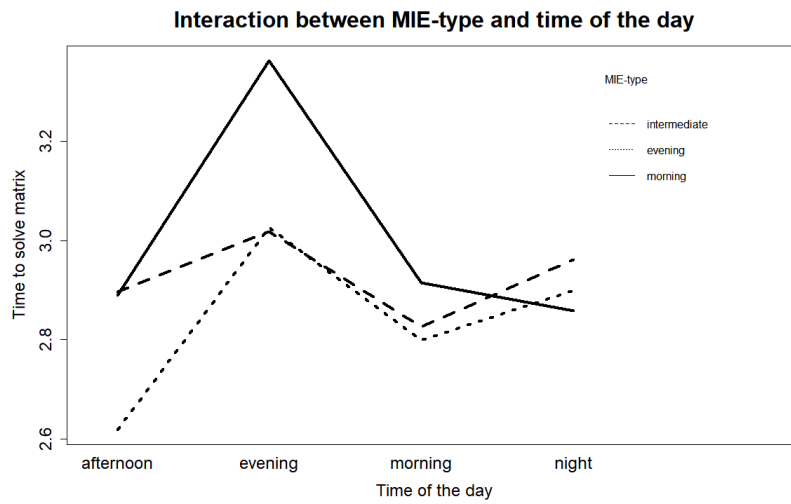
During the morning hours individuals needed significantly less time to solve a matrix than during the evening hours ($p = 0.04$, $M_{morning} = 2.83$, $M_{evening} = 3.12$). Moreover, more matrices were solved in the morning ($p = 0.02$, $M_{morning} = 27.14$, $M_{evening} = 25.03$) than in the evening. The performance was best in the morning, followed by the afternoon, the night hours and lastly by the evening hours.

We did not find any significant results based on the ME-type of the individuals. Morning type individuals are only slightly faster than evening type individuals ($M_{ME-M} = 2.91$, $M_{ME-E} = 2.94$), therefore, they solve slightly more matrices ($M_{ME-E} = 26.17$, $M_{ME-M} = 26.55$), and they have less mistakes ($M_{ME-E} = 0.40$, $M_{ME-M} = 0.22$).

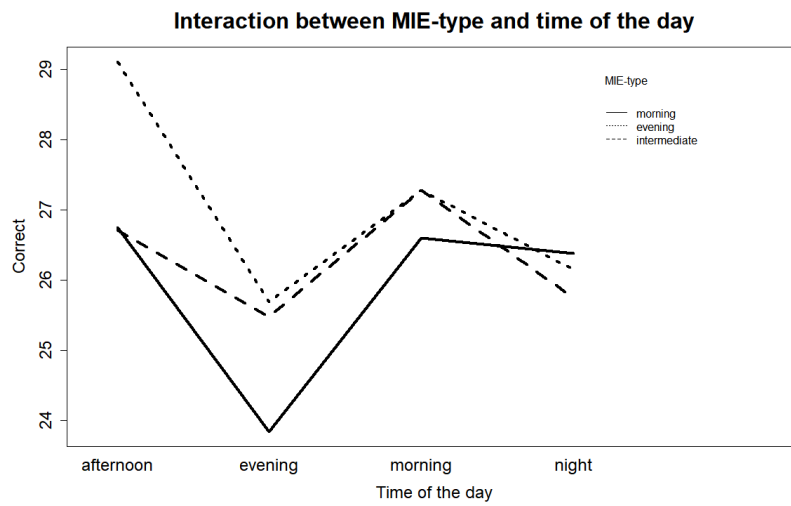
Concerning the MIE-type, we found that morning type individuals make less mistakes than intermediate type individuals ($p = 0.06$, $M_{MIE-M} = 0.12$, $M_{MIE-I} = 0.28$), and less mistakes than MIE-E type individuals ($p = 0.07$, $M_{MIE-E} = 0.60$). However, they have the highest reaction time ($M_{MIE-M} = 3.03$, $M_{MIE-E} = 2.85$, $M_{MIE-I} = 2.91$).

Next, we consider the interaction between the time of the day and the MIE-type of the individuals. We do not consider the interaction between the time of the day and the ME-type as no significant results were found. In Figure 6.12 are shown the three interaction plots.

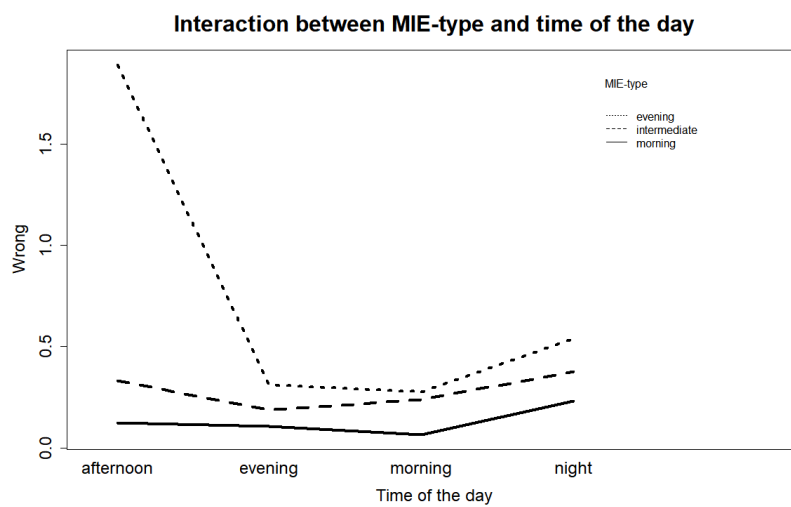
For the time needed to solve a matrix (Figure 6.12a), we can see that morning type individuals have the worst performance of all groups at all times of the day,



(A) Time to solve matrix



(B) Correct



(C) Wrong

FIGURE 6.12: Interaction plot for MIE-type and the interaction time for Integer Matrix Game

TABLE 6.10: Results for ME-M individuals during Integer Matrix Game

Measure	Morning time	Afternoon time	Evening time	Night time
Average time	$M = 2.780$	$M = 2.944$	$M = 3.208$	$M = 2.759$
Correct	$M = 27.61$	$M = 26.31$	$M = 24.86$	$M = 27.17$
Wrong	$M = 0.176$	$M = 0.291$	$M = 0.173$	$M = 0.217$

TABLE 6.11: Results for MIE-M individuals for Integer Matrix Game

Measure	Morning time	Afternoon time	Evening time	Night time
Average time	$M = 2.915$	$M = 2.889$	$M = 2.857$	$M = 2.857$
Correct	$M = 26.60$	$M = 26.75$	$M = 23.84$	$M = 26.38$
Wrong	$M = 0.066$	$M = 0.125$	$M = 0.105$	$M = 0.230$

except during the night, when they are the fastest group. Evening type individuals are the fastest in the morning, afternoon and during the evening. Intermediate type individuals, have a speed between that of the morning type and evening type individuals.

For the number of correct matrices (Figure 6.12b), the results are similar to that for the average time needed to solve a matrix. Morning type individuals solve the least matrices throughout the day, except at night, when they solve on average more matrices than the evening type and intermediate type individuals. Evening type individuals have the overall better performance throughout the day, with the best performance in the afternoon.

For the number of wrong matrices (Figure 6.12c), we can see that morning type individuals have the least number of wrong answers at all times of the day, followed by intermediate type individuals. Evening type individuals have the highest number of mistakes in the afternoon, but have the highest number of mistakes throughout the day.

These results show that while morning type individuals are not as fast as evening type or intermediate type individuals, they made the least amount of mistakes, showing that they tend to be more accurate.

Results for morning type individuals

First, we investigate the results of the ME-M type individuals (i.e., when dividing individuals into either morning or evening types). We did not find any significant results based on the time of the day.

In Table 6.10, we show the results of the ME-M type individuals. As can be seen, morning type individuals, are the fastest at night and in the morning, and solve the most matrices in the morning and at night. When taking into consideration all three performance measures, we can say that morning type individuals have the best performance in the morning, followed by the night hours. The worst performance is obtained in the evening.

No significant results were found for MIE-M type individuals either. Based on the results from Table 6.11, MIE-M type individuals have comparable performance in the morning, afternoon and at night. The worst performance was obtained during the evening hours.

TABLE 6.12: Results for MIE-E individuals for Integer Matrix Game

Measure	Morning time	Afternoon time	Evening time	Night time
Average time	$M = 2.798$	$M = 2.618$	$M = 3.026$	$M = 2.899$
Correct	$M = 27.27$	$M = 29.11$	$M = 25.68$	$M = 26.15$
Wrong	$M = 0.277$	$M = 1.888$	$M = 0.312$	$M = 0.538$

TABLE 6.13: Results for MIE-I individuals for Integer Matrix Game

Measure	Morning time	Afternoon time	Evening time	Night time
Average time	$M = 2.826$	$M = 2.896$	$M = 3.017$	$M = 2.961$
Correct	$M = 27.28$	$M = 26.71$	$M = 25.48$	$M = 25.75$
Wrong	$M = 0.237$	$M = 0.333$	$M = 0.185$	$M = 0.375$

Results for evening type individuals

For the ME-E type individuals, we found that the performance is dependent on the time of the day ($F(3, 130) = 3.45, p = 0.01$), with a better average time to solve a matrix in the afternoon compared to the evening ($p = 0.021, M_{evening} = 3.07, M_{afternoon} = 2.63$), and a better time in the afternoon than at night ($p = 0.05, M_{afternoon} = 2.63, M_{night} = 3.03$). Evening type individuals, have a comparable time to solve a matrix at night and in the evening, as well as in the morning and in the afternoon ($M_{morning} = 2.88$).

Considering the number of correctly solve matrices, we found that this result is dependent on the time of the day ($\chi^2 = 9.20, p = 0.026$), with the highest number of correctly solved matrices in the afternoon ($M_{afternoon} = 28.89$), and the least number of correctly solved matrices in the evening ($M_{evening} = 25.12$).

From these results, we can conclude that ME-E type individuals have the best performance in the afternoon, and the worst performance in the evening.

Based on the MIE-E type, we did not find any significant results. As shown in Table 6.12, evening type individuals are the fastest in the afternoon, but they also make the most number of mistakes at the same time.

Results for intermediate type individuals

No significant results were found for the intermediate type individuals. Based on the results in Table 6.13, we can see that intermediate type individuals have the best performance in the morning, followed by the afternoon. The worst performance was in the evening, followed by the night time.

Other results

Based on the BIG5 personality traits, we found that individuals with high levels for BIG5-A have made less mistakes than the individuals with low levels for BIG5-A ($\chi^2 = 7.07, p = 0.007$). Moreover, individuals with low levels for BIG5-O are faster ($\chi^2 = 6.62, p = 0.01$) and solve more matrices ($\chi^2 = 5.58, p = 0.01$) than individuals with high levels for BIG5-O.

Considering our second research question **RQ2: Does the BIG5-C personality trait influence how many times each game was played?**, we have found

TABLE 6.14: Distribution of the played Decimal Matrix games at different times of the day based on ME-type and MIE-type

Type	Decimal Matrix			
	Morning	Afternoon	Evening	Night
ME-type Morning	37	52	13	29
ME-type Evening	28	29	28	44
MIE-type Morning	16	18	11	20
MIE-type Intermediate	30	57	19	40
MIE-type Evening	19	6	11	13

that, individuals with high levels for BIG5-C played on average more games than individuals with low levels for BIG5-C ($M_{high} = 4.61$, $M_{low} = 4.41$, $\chi^2 = 0.01$, $p = 0.8$). However, this result is not significant.

6.3.3 Decimal Matrix

A total of 605 games were played by 117 participants. Out of these, we only selected the participants who played the game at least three times (i.e., in order to eliminate the influence of novelty from the results). Therefore, in the end, for the analysis we use 260 games played by 111 participants. The games were played at all hours of the day and by all types of individuals based on the results from MEQ and BIG5 (see also Table 6.14).

Task performance based on different factors

For the performance based on the time of the day, we did not find any significant results. Thus, the results at the different times of the day are not sufficiently different.

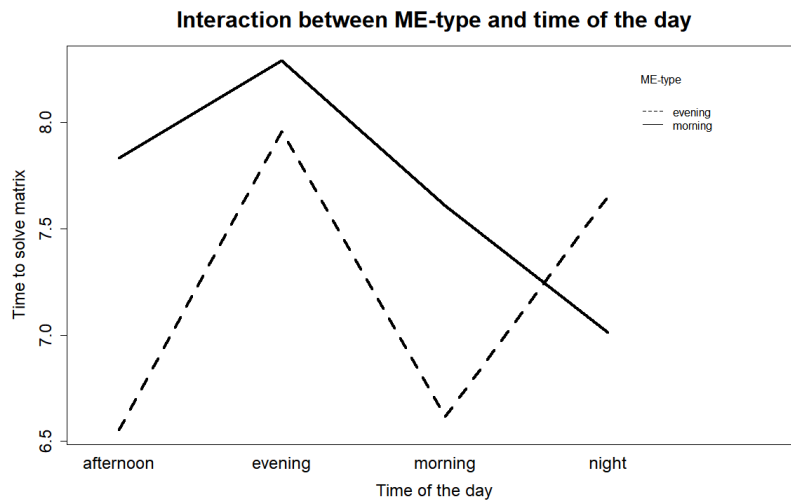
For the ME-type, we found that evening type individuals make more mistakes ($\chi^2 = 4.06$, $p = 0.04$) than morning type individuals ($M_{ME-E} = 1.37$, $M_{ME-M} = 0.88$). However, no significant results were found for the MIE-type. Therefore, considering that this task is more difficult than both previous tasks, we can conclude that for the Decimal Matrix Game, neither the time of the day, nor the ME-type or the MIE-type have an influence on the task performance.

As no significant results were found for MIE-type, we consider the interaction between the time of the day when a task is performed and the ME-type of the individuals. The results are shown in Figure 6.13.

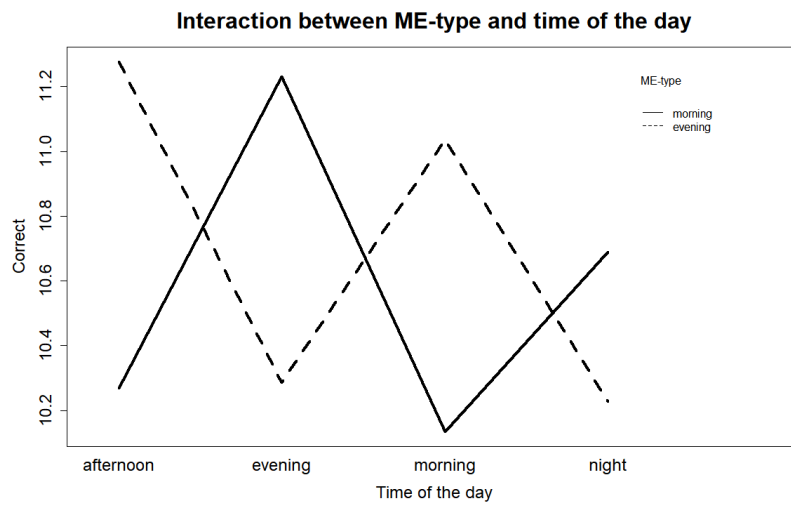
For the time needed to solve a matrix (Figure 6.13a), at all times of the day, except at night, evening type individuals are faster than morning type individuals. Also, for the evening type individuals, the reaction time changed depending on the time of the day, with better times in the morning and in the afternoon, than during the evening and at night.

For the number of correct matrices solved (Figure 6.13b), morning type individuals solve more matrices than evening type individuals in the evening and at night, while evening type individuals solve more matrices than morning type individuals in the morning and in the afternoon.

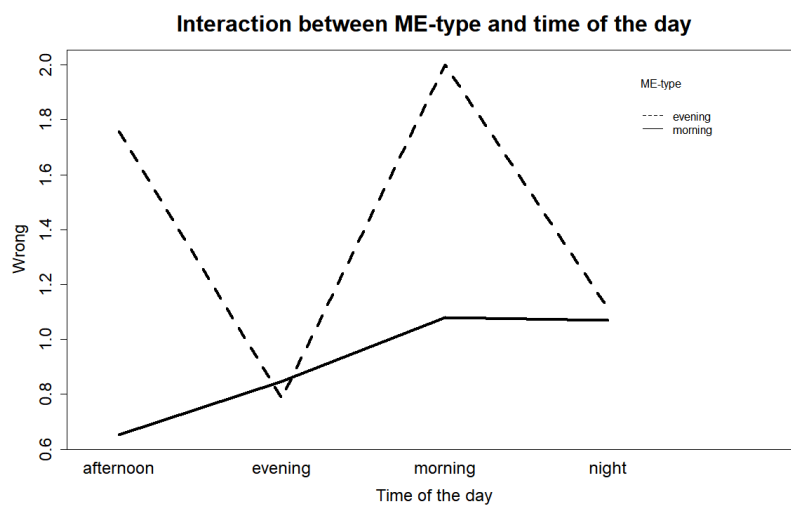
Based on the interaction plot shown in Figure 6.13c, morning type individuals make fewer mistakes than evening type individuals. There is a great variation in



(A) Time to solve matrix



(B) Correct



(C) Wrong

FIGURE 6.13: Interaction plot for ME-type and the interaction time for Decimal Matrix Game

TABLE 6.15: Results for ME-M individuals for Decimal Matrix Game

Measure	Morning time	Afternoon time	Evening time	Night time
Average time	$M = 7.607$	$M = 7.832$	$M = 8.289$	$M = 7.013$
Correct	$M = 10.13$	$M = 10.26$	$M = 11.23$	$M = 10.68$
Wrong	$M = 1.081$	$M = 0.653$	$M = 0.846$	$M = 1.068$

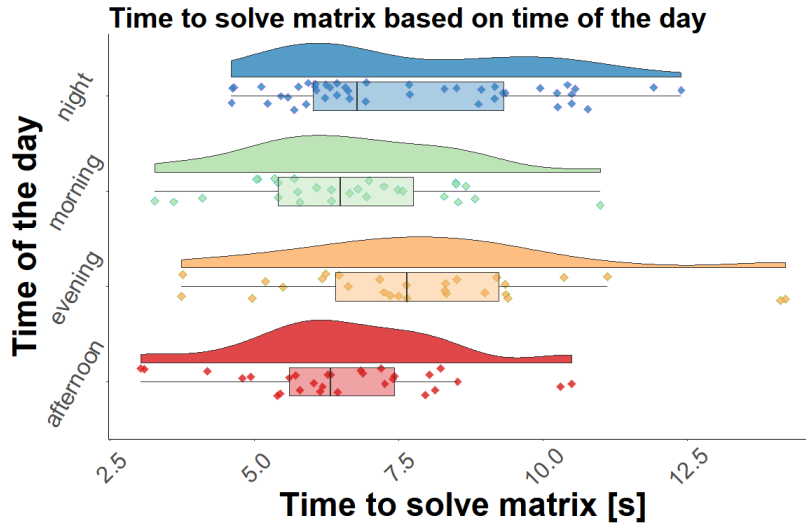


FIGURE 6.14: Time needed to solve a matrix based on the time of the day for ME-E type individuals during Decimal Matrix Game

performance for the evening type individuals, with more mistakes being made in the morning and in the afternoon, than in the evening and at night.

Results for morning type individuals

No significant results were found for ME-type individuals. However, based on the results presented in Table 6.15, ME-M type individuals have the best performance during the evening hours, followed by the the afternoon and the night hours. A similar trend was found for the MIE-M type individuals as well. However, it should be pointed out that these are just tendencies, as none of the factors have a significant influence on task performance.

Results for evening type individuals

For the MIE-E type individuals, the time of the day when the task was performed had no influence on the task performance. The most correct matrices were solved in the afternoon ($M_{afternoon} = 12.16$, $M_{morning} = 11.42$, $M_{evening} = 10.90$, $M_{night} = 11.07$). However, the second most mistakes were made also in the afternoon ($M_{morning} = 2.42$, $M_{afternoon} = 2.33$, $M_{evening} = 0.81$, $M_{night} = 0.84$). Therefore, while MIE-E type individuals do not solve the most matrices in the evening, they make the least mistakes in the evening.

Next, when we consider, the ME-E type individuals, the time of the day when the task was performed has an influence on the time needed to solve a matrix ($\chi^2 = 8.38$, $p = 0.03$). However, none of the pairwise comparisons show significant differences between the groups (see also Figure 6.14).

Results for intermediate type individuals

The results for the intermediate type individuals are quite similar between the different times of the day. On average, during each time of the day, approximately 10 matrices were correctly solve, and 1 mistake was made.

Other results

Other significant results were found for gender (correct answers - $\chi^2 = 5.63$, $p = 0.01$; wrong answers - $\chi^2 = 25.13$, $p < 0.0001$), BIG5-A level (reaction time - $\chi^2 = 9.45$, $p = 0.002$; correct answers - $\chi^2 = 10.78$, $p = 0.001$), BIG5-C level (reaction time - $\chi^2 = 5.79$, $p = 0.01$; correct answers - $\chi^2 = 9.27$, $p = 0.002$).

Considering the gender of the participants, males solved more matrices than females ($M_{male} = 11.07$, $M_{female} = 10.07$), and they also made fewer mistakes ($M_{male} = 0.74$, $M_{female} = 1.49$)

Individuals with low BIG5-A levels solved more matrices ($M_{low} = 11.27$, $M_{high} = 9.9$) and they were also faster ($M_{low} = 7.21$, $M_{high} = 7.63$) than individuals with high BIG5-A levels.

Individuals with low BIG5-C levels solved more matrices ($M_{low} = 11.04$, $M_{high} = 10.06$) and they were also faster ($M_{low} = 7.06$, $M_{high} = 7.78$) than individuals with high BIG5-C levels.

Considering our research question concerning the influence of BIG5-C level on how many games were played, we found that individuals with high BIG5-C levels played on average more games than individuals with low BIG5-C levels ($M_{high} = 4.77$, $M_{low} = 4.22$, $\chi^2 = 1.75$, $p = 0.18$).

6.4 Discussion and the message to take home

Our results show that the performance is dependent on the game, the time of the day when that game is being played, and the type of the individual who plays the game.

We found a greater influence of the time of the day and the type of individual (ME-type or MIE-type) for the Stroop game, followed by the Integer Matrix game. For the Decimal Matrix game we found the least influence of the time of the day and the ME-type or the MIE-type of the individuals on the game performance. We believe that this mostly related to the type of the games played. Stroop game is mostly an attention game, while both Integer and Decimal Matrix are games that require participants to perform mental calculations. The Decimal Matrix Game is the most difficult of the three games.

Moreover, we found that individuals have better performance in the morning than during the other hours of the day (for Stroop Game, Integer Matrix Game). For the Decimal Matrix Game, we found that evening type individuals have made more mistakes than morning type individuals.

Considering the ME-M and MIE-M type individuals, we found that for the Stroop game they are faster and more accurate in the morning than at other hours of the day. A similar result was found for the Integer Matrix Game, with the best performance in the morning and at night, and the worst performance in the evening. During the Decimal Matrix Game, morning type individuals had a better performance in the evening than at the other hours of the day.

Evening type individuals, if we consider the MIE-E type, they had the the best performance in the evening for the Stroop Game and that is when they made the

least mistakes during the Decimal Matrix Game. For the Integer Matrix Game, they had a better performance in the afternoon and morning.

For the intermediate type individuals, we found that they have a better performance in the morning for the Stroop Game and the Integer Matrix Game and no clear results were found for the Decimal Matrix Game.

As it can be seen, there is no clear time of the day when a specific type of individual has a better or worse performance. It depends on multiple factors (e.g., type of game). Furthermore, the study was in no way controlled. Morning type individuals had a good performance during the night but a bad performance during the evening. This could be attributed to sleep (e.g., a short nap in the evening), drinking caffeinated drinks, or other factors which we do not know about.

Therefore, we believe further study is required in order to better understand this relationship and how it can be applied to the field of social robotics. Just knowing the type of person it is currently, not sufficient, in order to determine when is the best time of the day a specific task should be performed.

6.5 Conclusion

In this Chapter, we presented the results of an online study in which we investigated the relationship between the M(I)E-type of 122 participants and the time of the day when they played three cognitive games and their influence on game performance. We have considered four times of the day (i.e., morning, afternoon, evening, and night). The participants had to play three cognitive games (Stroop Game, Integer Matrix Game, and Decimal Matrix Game), whenever they preferred. We considered for our analysis only the two best performances of the 128 participants. Our results show that there is a relationship between the M(I)E-type and the time of the day when the games are played. This information is of potential interest for social robotics, as this can lead to better performance depending on when the task is being performed and the type of task being performed. Moreover, we have found that individuals with high levels for BIG5 Conscientiousness personality trait, played on average more games than the individuals with low levels for BIG5-C. However, this result is not significant for any of the games played.

Further studies are needed with social robots as well, to better understand and use this relationship.

In the next chapter, we are going to investigate the influence of Empathy, Emotional Intelligence and Fight/Flight system in HRI.

6.6 My contribution

The main contribution of this chapter stands in the development of the online platform and in the large scale non controlled study. Furthermore, our study is one of the few studies that investigate the intermediate type individuals on the morningness-eveningness scale) as well. Moreover, three cognitive tasks were investigated in this study: the Stroop Game, Integer, and Matrix Task.

Chapter 7

RQ3: Influence of Empathy, Emotional Intelligence and Fight/Flight system in HRI

7.1	Introduction	81
7.2	Experimental design	82
7.2.1	Robotic platform and sensors	82
7.2.2	Questionnaires	82
7.2.3	Participants	83
7.2.4	Scenario	83
	Robot behavior	84
7.2.5	Hypothesis	86
7.3	Data extraction and analysis	86
7.3.1	GSR	86
7.3.2	Facial Temperature variation	87
7.4	Results	87
7.4.1	Panas questionnaire	87
7.4.2	GSR	90
7.4.3	Facial Temperature variation	93
7.5	Discussion	94
7.6	Conclusion	95
7.7	Contribution	95

7.1 Introduction

In their everyday lives humans interact with each other. They might notice the others around them, or they may choose to ignore them. The way an individual reacts to another individuals' actions defines the concept of empathy (Davis, 1983). Empathy is an important construct especially in social interactions.

Can the concept of empathy be used in the interaction between a human and a robot? One possible answer to this question was given by the authors of (Tapus et al., 2007b). They state that even though a machine cannot feel empathy, it could display a behavior that is empathetic. In order to accomplish this, the authors of (Tapus et al., 2007b) state that "a robotic system should be capable of recognizing the user's emotional state, communicating with people, displaying emotion". Research has shown that the emotional state of an individual has an effect on different physiological parameters (Picard et al., 2001; Ioannou et al., 2014b). It was also shown that there are multiple physiological parameters that can be used to determine an

individual's emotional state (e.g., heart rate, facial temperature variation (Ioannou et al., 2014b), galvanic skin response (GSR) (Lisetti et al., 2004), blinking).

In this chapter, we are focusing on the understanding of the emotional state of an individual when it is interacting with a robot around a task, in an unforeseen situation. More specifically, while doing a certain task (i.e., playing the Jenga game), the participants are abruptly interrupted either by knocking over their tower, or by telling them that they have to stop. Of interest for this research, is how different physiological parameters vary (i.e., GSR, facial temperature variation) based on who the participants are interacting with (either a robot or a human), if their tower is knocked over or not, their empathy level, their emotional intelligence, and their personality.

7.2 Experimental design

7.2.1 Robotic platform and sensors

For this experiment the Pepper robot (see Chapter 3.3) was used. Data from two sensors was recorded: thermal camera and a GSR sensor. The thermal camera was placed in front of the participants, while the GSR sensor was placed on the ring and middle finger of the left hand of each participant.

7.2.2 Questionnaires

In Chapter 4.5, we presented the main psychological questionnaires used in this thesis. However, for this work, we needed to include other questionnaires that can help us define the user profile for this experimental task. Therefore, below, we present the questionnaires that are specific for this experiment.

The participants had to fill out the following questionnaires:

Toronto Empathy Questionnaire (TEQ) (Spreng et al., 2009) is a self-report instrument to measure the empathy level. In general, by empathy, it is meant the consequences of accurately perceiving how another individual is feeling. The more empathic an individual, the better it understands what the person it interacts with is feeling, and the more appropriately it can react. The questionnaire consists of 16 questions on a 5 point Likert scale (0 - "Never", 4 - "Always").

Trait Emotional Intelligence Questionnaire (TEIQ) (Petrides, 2009) measure the emotional intelligence of an individual and it considers it as a personality trait. The model is based on 4 traits: emotionality, sociability, well-being, and self-control. For example, an individual with a high self-control is capable of better controlling its impulses than an individual with low self-control.

The Reinforcement Sensitivity Theory Personality questionnaire (RST-PQ), presented in more details in Chapter 4.5.

Positive And Negative Affect Schedule (PANAS) (Watson et al., 1988) was developed to measure both positive and negative affect. It consists of 20 items, each measured on a scale from 1 ("Not at all") to 5 ("Very much"). High positive affect is characterized by high energy, full concentration, while negative affect is characterized by distress and unpleasurable engagements.

Post-questionnaire a custom designed questionnaire was given to the participants in order to assess how they perceived the experimenter (e.g., friendly, motivating, empathetic), the task (i.e., stressful, difficult) and if the breathing exercise helped them relax or not. Each question was measured on a scale from 1 ("Strongly disagree") to 5 ("Strongly agree").

TABLE 7.1: Participants distribution based on questionnaires

RST-PQ						
Category	FFFS	BIS	BAS RI	BAS RR	BAS GDP	BAS I
low	12	13	12	12	12	13
high	11	10	11	11	11	10

TEIQ				
Category	Well being	Sociability	Emotionality	Self Control
low	15	13	14	13
high	8	10	9	10

TEQ	
Category	Empathy
low	13
high	10

Robotics knowledge				
Not at all	A little	Somewhat	Much	Very much
1	4	6	5	7

For all traits of the TEQ, TEIQ, and RST-PQ, the scores are considered either low or high, by using a threshold equal to the median of the participants that took part in this study.

7.2.3 Participants

A total of 23 participants (i.e., 4 female and 19 male, mean age of 27.48, $SD = 5.76$) agreed to take part in this experiment. Most of the participants have a technical background (21 out of 23), one has social and cognitive sciences background, and one has non technical background. Before the experiment, each participant's knowledge of robotics was measured on a scale from 1 ("Not at all") to 5 ("Very much") (see Table 7.1).

Table 7.1 summarizes the results of the participants to the questionnaires presented in Section 7.2.2. Based on these results, the following questionnaire traits were considered for further analysis: all traits of RST, empathy, sociability and self control (from TEIQ).

7.2.4 Scenario

For this study, the participants had to play the Jenga game (see Figure 7.1a). Given a tower of 54 pieces of wooden blocks (3 blocks per layer), the purpose of the game is for the user to extract a block from any of the layers of the tower and place it on the top of the tower. The game ends when the tower falls. The participants sat at a coffee table, where the tower was already built. They were instructed to play the game and to do their best to build the tallest tower. Each participant interacted with an experimenter (either a robot or a human). During the game, the experimenter would periodically give encouragements and cheer the participant.

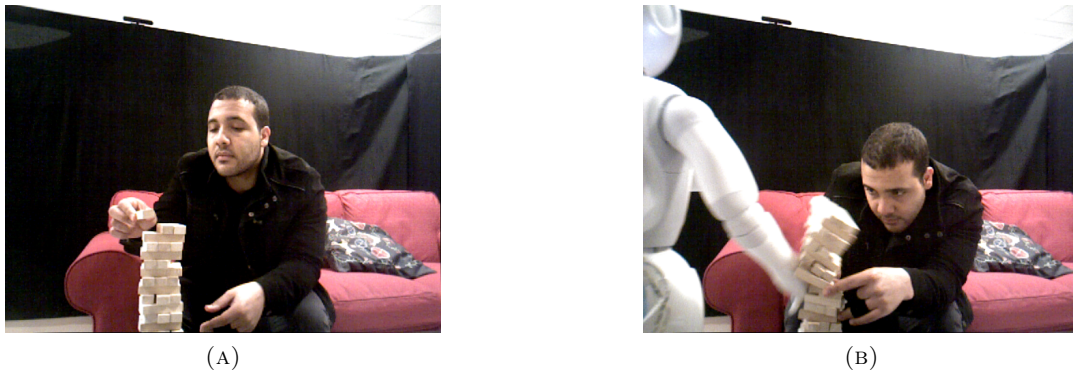


FIGURE 7.1: (a) Participant placing a Jenga piece; (b) Experimenter knocks over the tower

Once seated at the table, and before the experiment started, there was a 5 minute relaxation period, in which each participant had to follow a breathing exercise. The purpose of this relaxation period was for the participants to start playing the game from a calmed state. Once the relaxation was done, the experimenter explained to the participants what they had to do. After the instructions, each participant filled the PANAS questionnaire and the interaction with the experimenter started. First, there was a short dialogue between the experimenter and the participant. The purpose of this, was so that the participant familiarizes itself with the experimenter. Both the human and the robot experimenters had the same behavior, which is explained in Section 7.2.4.

After approximately two minutes of playing the game, the experimenter approaches the participant. In some of the cases it will bump into the table and the tower falls (see Figure 7.1b), while in some other cases it just approaches the participant and it informs him/her that the game is over. When the experiment was finished, each participant had to fill a new PANAS questionnaire.

In our experiment, four conditions were developed. More specifically:

Condition C1: Interaction with Robot experimenter with a strong bump that makes the tower fall

Condition C2: Interaction with Robot experimenter with a bump not strong enough to make the tower fall

Condition C3: Interaction with Human experimenter with a strong bump that makes the tower fall

Condition C4: Interaction with Human experimenter with a bump not strong enough to make the tower fall

The participants were assigned a certain condition based on their empathy levels. In the end, the distribution shown in Table 7.2 was obtained.

Next, the behavior of the robot is going to be presented.

Robot behavior

For the dialog between the robot and the participant the Text To Speech (TTS) and the Automatic Speech Recognition (ASR) modules, provided by the NaoQI Framework were used. The TTS was used to generate the speech of the robot, while the

TABLE 7.2: Participants distribution based on the conditions

Condition	Low empathy	High empathy
C1	5	2
C2	3	3
C3	2	2
C4	3	3

TABLE 7.3: Angles of arm joints of the robot when knocking down the tower

	Elbow Yaw	Hand	Shoulder Pitch	Shoulder Roll	Wrist Yaw
Right	70.5°	0.61°	36.8°	-40.5°	43.2°
Left	-70.5°	0.61°	36.8°	40.5°	-43.2°

ASR was utilized to recognize the instruction of the participants to "start" the game. The body gesture of the robot was designed using Choregraphe.

The robot started the interaction with a pre-programmed speech with pauses between phrases. The duration of the pauses used are as follows: "Hello", wait (1 sec), "now you can proceed to complete the questionnaire on the table", wait (3 secs), "When you finish it please let me know by saying **start**", wait (until participant said "start"), "How are you?", wait (6 secs), "I'm fine, my name is Pepper, and I am here to stand by you during the game", wait (6 secs), "What is your name?", wait (8 secs), "It is very nice to meet you, what do you think about our game Jenga?", wait (7 secs), "Ok. This game is funny, and even more when somebody is motivating you. I hope you do very well. You can start now", wait (7 secs).

Every twenty seconds, the robot moved autonomously backwards and forwards (it started by first moving backwards) covering a distance of 50 cm. Then, after ten seconds it said a phrase to cheer up the participant while opening the arms (as shown in Figure 7.2). The angles of the arm joints of the robot used for the postures are shown in Table 7.3. Some examples of the encouraging phrases said by the robot are: "It seems you are a good player", "That was a great movement", "You are doing it very well".

Once 110 seconds past since the participant said "start", if in condition C1 (making the tower fall), the robot moved forward with open arms while saying "look, you are awesome". The robot continued moving until bumping into the table with its base and knocking down the tower with its arms (see Figure 7.1b). Then, the robot said "Oh! I am so sorry!", wait (5 secs), "Ok. I think you cannot keep playing this game, I'm sorry", wait (10 secs), "Please fill the questionnaire on the table". In condition C2, the robot just informed the participant that the time was over, and asked him/her to fill the PANAS questionnaire that was on the table.

For conditions C3 and C4, with the human experimenter, the experimenter behaved in the same way as the robot. It followed the same steps and questions with the dialog, and it encouraged the participant every 20 seconds. When the 110 seconds were elapsed, the experimenter either bump into the table, or not, depending on the condition.

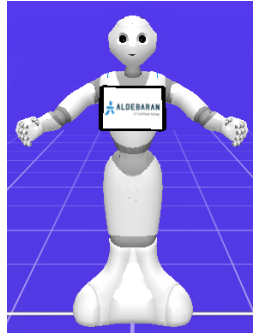


FIGURE 7.2: Robot posture when cheering up the participants and knocking down the tower

7.2.5 Hypothesis

Based on the information presented so far the following hypothesis was developed:

H. Participants in conditions C1 and C3 (where the tower fell) will show a greater physiological and mood change response, than participants in conditions C2 and C4 (where the tower does not fall) (as measured by GSR event based analysis and PANAS questionnaire).

7.3 Data extraction and analysis

In this section, the methods used to extract and analyze the data are described.

7.3.1 GSR

Based on the information presented in Chapter 4.4, it is known that there are two types of analyses that can be performed: analysis on the entire interaction and event based analysis. For the entire interaction analysis the accumulative GSR (AccGSR) (Nourbakhsh et al., 2012) was extracted, while for the event based analysis four parameters were considered (Setz et al., 2010): latency time, rise time, amplitude and recovery time. The event considered is the moment when the experimenter (either robot or human) bumped into the table (for conditions C1 and C3). In case when the tower did not fall (C2 and C4), the event is considered when the experimenter informs the participant that the game is over. A typical recorded GSR signal is shown in Figure 7.3. The upper part of the figure presents the variation of the GSR signal and the detected peaks, while in the lower part of the figure is shown the GSR signal corresponding to the bump event. In both figures, the vertical line, represents the time at which the bump occurred.

The **latency time** (measured in seconds) was computed as the total time it took for the signal to increase with at least 10 % compared to the level at the bump time. The **amplitude** represents the difference between the maximum value of the signal and the level at the bump time. The **rise time** (measured in seconds) represents the time it took the signal to increase from the value at the latency time to the time when the amplitude is reached. The **recovery time** (measured in seconds) is computed as the time difference between the time when the signal reaches a level of 63 % of the amplitude and the time when the maximum value is reached.

The distributions of the four measures are shown in Figure 7.4. A visual analysis of the data suggests that the four measures contain outliers which need to be dealt with. According to Tukey's method, a mild outlier lies between 1.5 times and 3 times

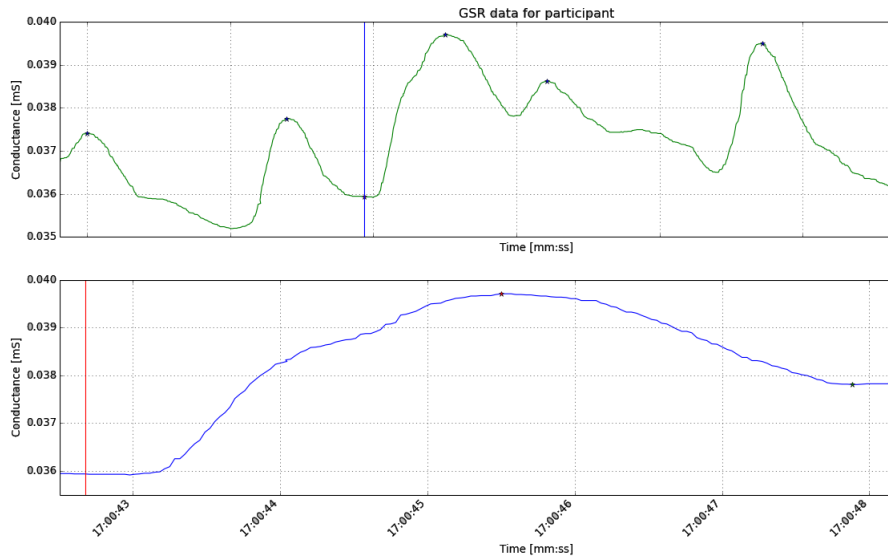


FIGURE 7.3: Typical data for event based analysis. The vertical lines represent the bump events.

the interquartile range below the first quartile or above the third quartile. An extreme outlier tends to lie more than 3 times the interquartile range below the first quartile or above the third quartile.

For the latency time, the data contains one mild outlier and one extreme outlier. The amplitude data does not contain any outliers. The rise time data contains two mild outliers and one extreme outlier, and finally the recovery time data contains one mild outlier. All the extreme outliers have been replaced with the average value of the data.

7.3.2 Facial Temperature variation

The temperature variation across different regions of interest (ROI) provides good insight into the current internal state of an individual. Therefore, the thermal data was used to extract these temperature variations. The ROIs considered are: the entire face, the forehead, the left, and right periorbital regions, the nose, the perinasal region, the left, and right cheek, and the chin region. A description of how these regions are defined is presented in Chapter 4.2 and can also be seen in Figure 7.5. For analysis the rate of change of the temperature (i.e., slope) after the bump is considered. All slopes which are not significant (p value less than 0.05) were not considered for the analysis.

7.4 Results

Next, the main results that were obtained are presented.

7.4.1 Panas questionnaire

As previously mentioned, the participants completed a PANAS (Watson et al., 1988) questionnaire before, and after the experiment. Positive and negative mood differences were computed between the two questionnaires (mood after the experiment -

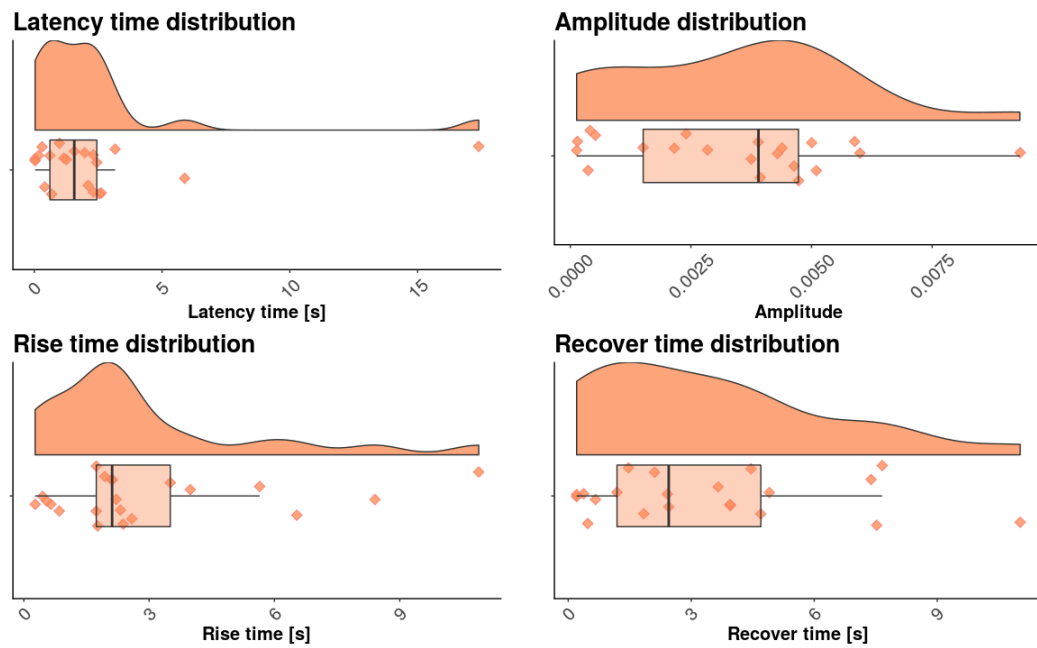


FIGURE 7.4: Distribution of the four GSR event based parameters: latency time, amplitude, rise time, recovery time

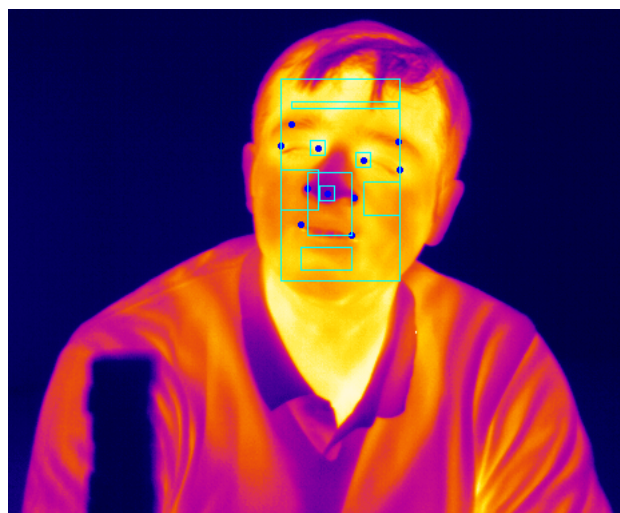


FIGURE 7.5: Example of ROIs

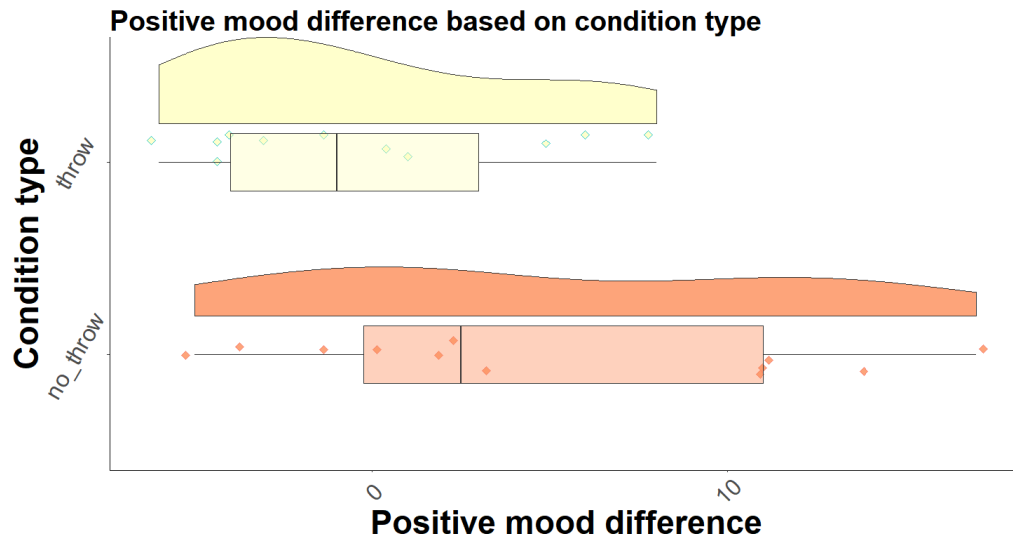


FIGURE 7.6: Positive mood difference based on the condition type (throw/no throw)

mood before the experiment). Both parameters, positive difference and negative difference, showed a normal distribution ($p = 0.06$; $p = 0.22$). Therefore, an ANOVA analysis could be applied on the data. Statistical analysis yielded the following significant results.

The condition type parameter approaches significance ($F(1, 21) = 4.09, p = 0.056$), with participants in the no fall conditions (C2 and C4) showing a greater positive mood difference ($M=5.08$) than participants in the fall conditions (C1 and C3) ($M=-0.18$) (see Figure 7.6). As a result, our **Hypothesis is partially validated**.

Participants with high sociability scores showed a significantly higher positive mood difference ($M=6.7$) than participants with low sociability scores ($M=-0.61$) ($F(1, 21) = 9.436, p = 0.0057$). On the other hand, participants with high scores in self control had a lowered negative mood after the interaction (mood difference of $M=-2.3$) than the participants with low self control scores (mood difference of $M=0.92, F(1, 21) = 6.32, p = 0.02$).

Considering the Fight/Flight/Freeze system, individuals with high RST-FFFS scores showed a higher positive mood difference ($M=6.36$) than participants with low RST-FFFS scores ($M=-0.91, F(1, 21) = 9.514, p = 0.0056$). The same relationship was found for the RST-RR scores as well. More specifically, participants with high reward reactivity scores showed a higher positive mood change ($M=5.63$) than participants with low scores ($M=-0.25, F(1, 21) = 5.376, p = 0.03$).

As shown in Section 7.2.2, each of the following statements had to be rated by the participants on a scale from 1 ("Strongly disagree") to 5 ("Strongly agree"):

(a) **“The experimenter was polite”.**

Considering this statement ($F(1, 21) = 14.4, p = 0.001$), participants who gave a rating of 4 (“Agree”, $N=9$) showed a significantly lowered negative mood ($M=2.1$) than the participants who gave a rating of 5 (“Strongly Agree”, $N=13, M=-2.46, p=0.0043$). More exactly, participants who gave a rating of 5 had a lower negative mood after the interaction than before the interaction.

(b) **“The experimenter was friendly”.**

Considering this statement ($F(1, 21) = 7.79, p = 0.01$), participants who gave a rating of 4 (“Agree”, $N=10$) showed a significantly lowered negative mood ($M=1.2$) than the participants who gave a rating of 5 (“Strongly Agree”, $N=9, M=-3, p=0.012$).

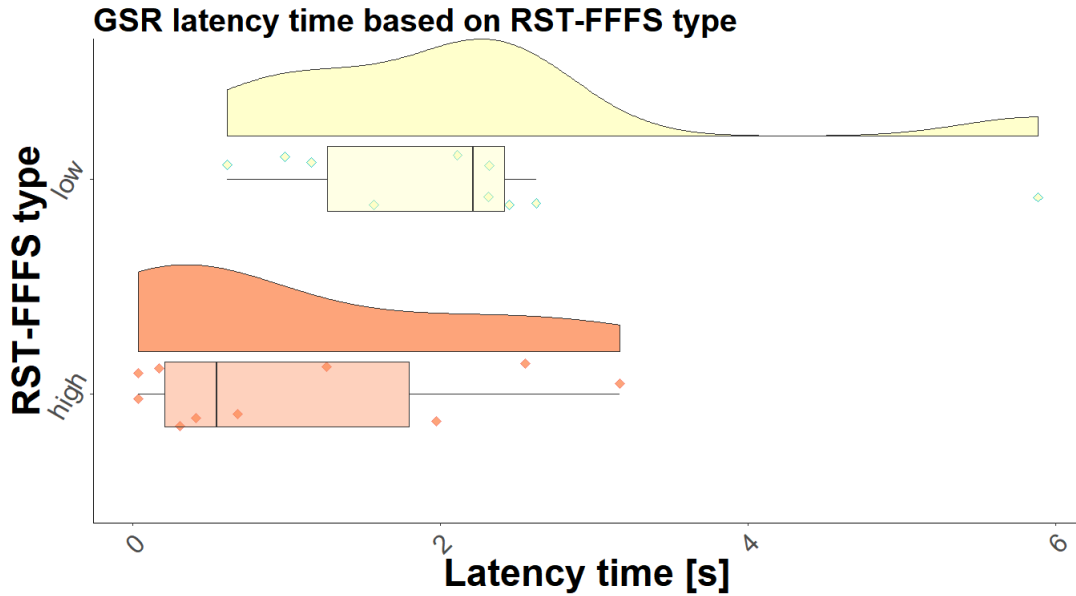


FIGURE 7.7: Latency based on RST-FFFS

More exactly, participants who gave a rating of 5 had a lower negative mood after the interaction than before the interaction.

(c) “The experimenter was helpful”.

Considering this statement ($F(1, 21) = 8.9, p = 0.007$), participants who gave a rating of 2 (“Disagree”, $N=6$) showed a significantly lower negative mood ($M=2.16$) than the participants who gave a rating of 5 (“Strongly Agree”, $N=5, M=-4.2, p=0.012$). More specifically, the participants who gave a rating of 5 had a significantly lower negative mood after the interaction than before the interaction compared with the participants who gave a rating of 2.

7.4.2 GSR

The GSR analysis was performed on the latency time, rise time, amplitude, and recovery time, for the event based analysis, and the AccGSR for the relaxation period, the entire interaction, and after the bump into the table.

Latency time:

A Shapiro-Wilk normality test showed that the latency was not normally distributed ($p < 0.05$). Therefore, a Kruskal-Wallis test was applied on the latency variable by using the different variables as factors. A significant result was found for the RST-FFFS ($\chi^2 = 3.89, p = 0.048$). Individuals with high RST-FFFS scores had a significantly lower latency time than the individuals with low RST-FFFS scores (see Figure 7.7). Based on this result, we can state that our **Hypothesis** is partially confirmed. More specifically, individuals with high scores for the fight/flight/freeze system need less time ($M=1.05$) to start to react than individuals with low scores ($M=2.21$). Meaning that they are always prepared to react in unforeseen situations.

Amplitude:

As the amplitude parameter shows a normal distribution ($p > 0.05$), an Anova analysis was performed. A result that approaches significance was found for the condition type ($F(1, 19) = 3.87, p = 0.06$). Individuals in the throw conditions (C1 and C3) showed higher amplitude levels ($M=0.0043$) for the GSR signal than

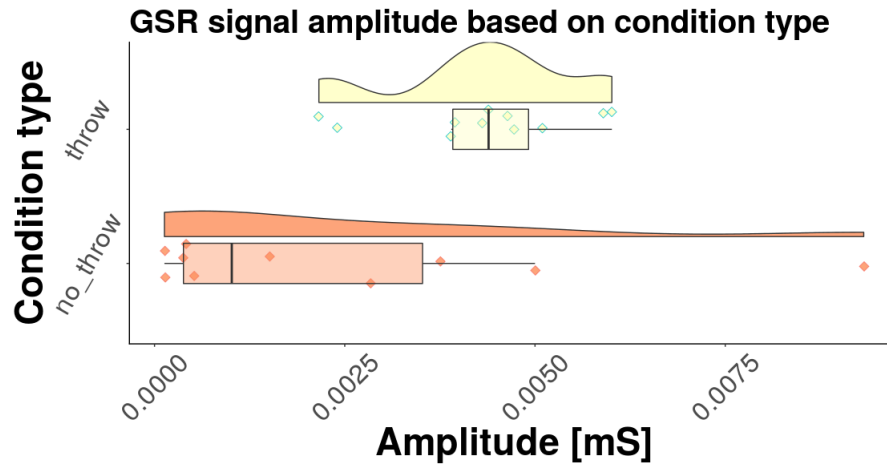


FIGURE 7.8: Amplitude based on condition type

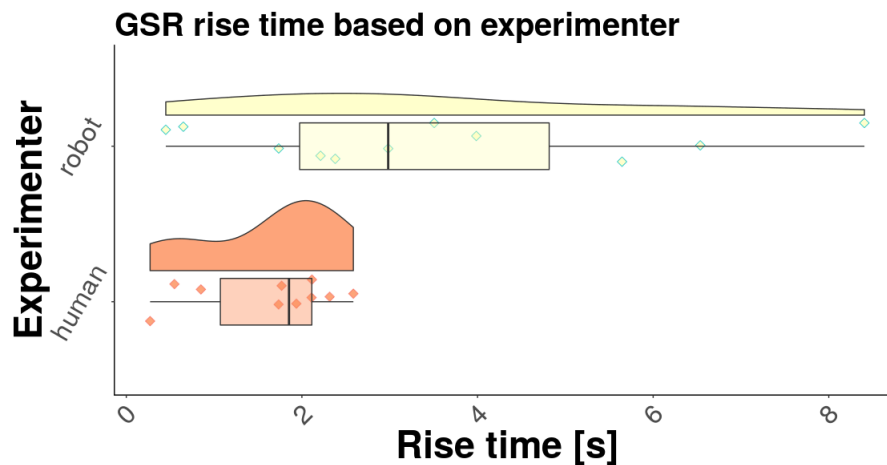


FIGURE 7.9: Rise time based on experimenter

the individuals in the no-throw conditions (C2 and C4) ($M=0.0024$). A graphical representation of this result can be found in Figure 7.8.

Rise time:

The rise time measure does not show a normal distribution ($p < 0.05$), as a result a Kruskal-Wallis analysis was performed. The significant factors found are the condition type and the experimenter.

Considering the condition type ($\chi^2 = 4.46, p = 0.03$), individuals in the no-throw conditions (C2 and C4) had a significantly lower rise time ($M=1.61$) than the individuals in the throw conditions (C1 and C3) ($M=3.51$). While interacting with the robot, the participants had a longer rise time ($M=3.49$) than the participants interacting with the human experimenter ($M=1.62$), ($\chi^2 = 4.17, p = 0.04$), see Figure 7.9. No significant interaction of the two factors was found.

Recovery time:

The recovery time parameter did not show a normal distribution ($p < 0.05$), therefore a Kruskal-Wallis analysis was performed. Significant results were found for the condition, condition type, and experimenter.

For the condition factor ($\chi^2 = 12.4, p = 0.006$), see Figure 7.10, differences were

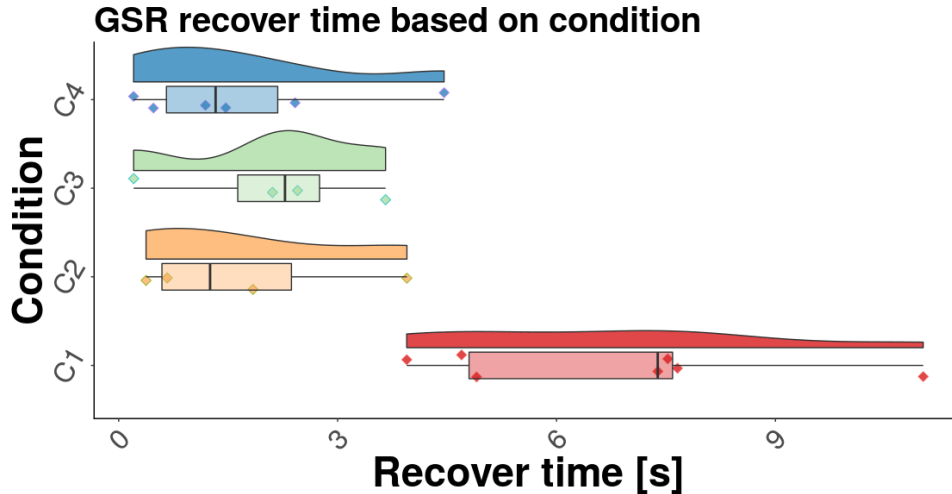


FIGURE 7.10: Recovery time based on condition

TABLE 7.4: Results for the GSR parameters

Measure	Factor	Test	Statistic (F or χ^2)	p-value
Amplitude	condition type	Anova	3.87	0.06
Latency	RST-FFFS	Kruskal-Wallis	3.89	0.048
Rise time	condition type	Kruskal-Wallis	4.46	0.03
Rise time	experimenter	Kruskal-Wallis	4.17	0.04
Recovery time	condition	Kruskal-Wallis	12.4	0.006
Recovery time	condition type	Kruskal-Wallis	6.78	0.0091
Recovery time	experimenter	Kruskal-Wallis	5.4	0.02

found between condition C1 and condition C3 ($p = 0.036$). Participants in condition C1 had a significantly higher recovery time ($M=6.73$) than the participants in condition C3 ($M=2.1$).

Significant results were found for the condition type ($\chi^2 = 6.78, p = 0.0091$), with the participants in the throw conditions showing a significantly higher recovery time ($M=5.05$) than the participants in the no throw conditions ($M=1.7$).

When considering the experimenter ($\chi^2 = 5.4, p = 0.02$), the participants who interacted with the robot showed a higher recovery time ($M=4.9$) than the participants that interacted with the human experimenter ($M=1.86$).

An overview of the GSR results is presented in Table 7.5

Considering all four event based GSR parameters (i.e., latency time, amplitude, rise time, and recovery time), participants in the throw conditions showed an increased rise times, higher amplitude, and higher recovery time, than the participants in the no throw conditions. Taking into account the result of the PANAS questionnaire, it can be stated that our **hypothesis (Participants in conditions C1 and C3 (where the tower fell) will show a greater physiological and mood change response, than participants in conditions C2 and C4 (where the tower does not fall) (as measured by GSR event based analysis and PANAS questionnaire)) is validated.**

TABLE 7.5: Results for the AccGSR

Phase	Factor	Test	Statistic (F or χ^2)	p-value
Relaxation	Well being	Kruskal-Wallis	4.82	0.02
Bump	Condition	Kruskal-Wallis	9.68	0.02
Bump	Condition type	Kruskal-Wallis	5.76	0.01
Bump	BAS-I	Kruskal-Wallis	3.93	0.04
Bump	Empathy	Kruskal-Wallis	4.18	0.04

Regarding the AccGSR parameter, the significant results are shown in Table 7.5.

During the relaxation phase, individuals with high scores for the well-being emotional intelligence trait had a lower value for the AccGSR ($M=1.09$) than the participants with low scores for well being ($M=1.73$). No significant results were found for the AccGSR throughout the experiment. In the throw conditions the AccGSR parameter is higher ($M=1.85$) than during the no-throw conditions ($M=1.6$), meaning that during the throw conditions the participants were more stressed. However, this result is not significant, it is just a trend that we noticed.

The ‘‘Bump’’ phase means the interaction that took place after the experimenter knocked over the Jenga tower. Looking at the average AccGSR values in each of the four conditions (C1 - 0.43, C2 - 0.19, C3 - 0.26, C4 - 0.21), a significant difference was found for conditions C1-C2 ($p = 0.028$). The participants that interacted with the robot that knocked over the tower had significantly higher AccGSR values than the participants that interacted with the robot that did not knock over their tower. During conditions C2 and C4 (when the tower did not fall) the AccGSR values were significantly lower ($M=0.2$) than during conditions C1 and C3 ($M=0.32$) (when the tower fell). As the AccGSR is a good measure for stress, we can conclude that the participants were more stressed after the experimenter knocked over their tower, than when the experimenter did not knock over their Jenga tower.

Another significant result was found for impulsivity, individuals with low impulsivity levels had higher AccGSR values ($M=0.32$) than individuals with high impulsivity levels ($M=0.22$).

Individuals with high empathy levels had lower AccGSR values ($M=0.23$) than individuals with low empathy levels ($M=0.32$).

7.4.3 Facial Temperature variation

For the temperature variation, the rate of change of the temperature variation was analysed. The temperatures were extracted a few seconds before the event (i.e., the moment when the experimenter bumped into the table) and up to 10 seconds after the event.

The temperature variation in the chin region is dependent on the condition type (Kruskal-Wallis analysis, $\chi^2 = 5.13, p = 0.023$). Individuals in the throw conditions had significantly higher temperature changes ($M=0.013$) than the individuals in the no-throw conditions ($M=-0.014$).

When considering the RST-FFFS as factor, four regions of interest showed either significant or almost significant results: the forehead ($\chi^2 = 3.91, p = 0.047$), the left periorbital region ($\chi^2 = 4.05, p = 0.044$), the right periorbital region ($\chi^2 = 3.72, p = 0.053$) and the entire face region ($F(1, 20) = 6.06, p = 0.0043$). For all

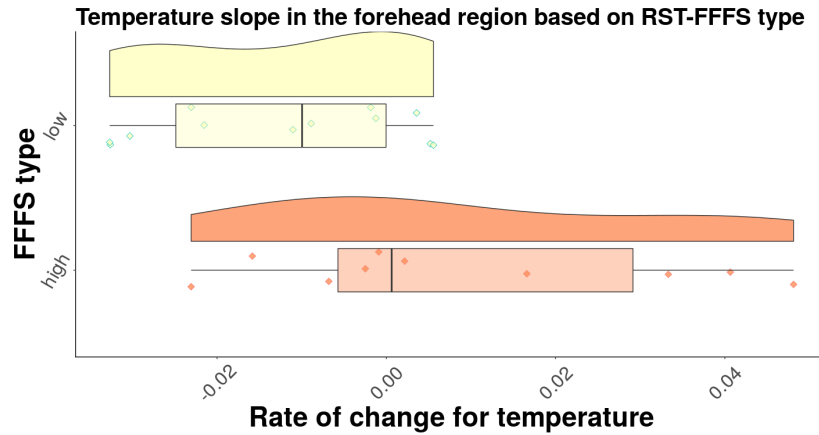


FIGURE 7.11: Forehead temperature variation based on RST-FFFS type

regions, individuals with high RST-FFFS scores had a significantly greater temperature variation than the participants with low scores. Furthermore, participants with high scores showed a temperature increase, while participants with low scores showed a temperature decrease.

The temperature variation in the left and right periorbital regions is dependent on the RST-BIS type ($\chi^2 = 5.0, p = 0.025$ for the left region, and $\chi^2 = 7.71, p = 0.0054$ for the right region, respectively). For both regions, individuals with high RST-BIS scores showed a temperature decrease, while the individuals with low RST-BIS scores showed a temperature decrease (for the left periorbital region) or a temperature increase (for the right periorbital region).

Another aspect investigated was the difference between the average temperature during the bump and the average temperature during the relaxation phase. A significant result was found for the forehead region depending on RST-I type ($F(1, 21) = 4.69, p = 0.04$). For all individuals, the forehead temperature decreased after the bump with a greater decrease for the individuals with high scores for RST-I ($M = -1.02$) compared to the individuals with low RST-I scores ($M = -0.3$).

7.5 Discussion

In this study the main emphasis was put on the GSR physiological parameter, as it was proved in the literature to be a good indicator of an individual's arousal level (Nebylitsyn et al., 1972), or of its emotional state (Lisetti et al., 2004). For the GSR data significant results were found for the event based analysis parameters. Significant differences were found between conditions, condition types (either throwing or not throwing the tower) and experimenter. As the latency time, rise time, amplitude and the recovery time were extracted for a specific event, it should not be surprising that no results were found for any of the psychological questionnaires.

As shown in Section 7.2.2, the FFFS is responsible with how individuals react in "get me out of here" situations. As a result, it is expected of individuals with high scores to be better prepared to flee from a situation than an individual with low scores. Therefore, as shown in (Ioannou et al., 2014a), this situation should be characterized by a greater electrodermal activity. When looking at the temperature variation, significant results were found between how the temperature varies in the forehead, the entire face, and the periorbital regions and the FFFS. According to (Ioannou et al., 2014a), individuals with high scores should have a lower increase in

temperature than individuals with low scores. Considering that the temperature was extracted a little before and after the event, our results can be explained by the fact that individuals with high scores were ready to fight or flee (therefore they showed an increase in temperature), while individuals with low scores froze (they showed a decrease in temperature).

The main limitations of our work consists in the relatively small number of participants (i.e., 23 participants) that took part in our study. More investigation is needed in order to confirm the universality of our results.

7.6 Conclusion

In this chapter, a between participants study was presented that was carried out with 23 participants, with the purpose of finding out how an individual would react a situation where an experimenter accidentally knocks over a Jenga tower. Four conditions have been developed, two with a robot experimenter and two with a human experimenter, in which the experimenter (either robot or human) either knocked over the tower of the participant or not. Different psychological questionnaires were used to measure the participants empathy level, emotional intelligence, and personality (i.e., TEQ, TEIQ, RST-PQ). Also, different physiological parameters (GSR, temperature variation) were extracted to better understand the reactions of the participants. **The results confirmed all our hypotheses.**

Most of our results are based on the GSR parameters. We found evidence that the event based analysis parameters (i.e., latency time, rise time, amplitude, recovery time) are dependent on the condition, and if the experimenter makes the tower fall or not and who the experimenter is (i.e., robot or human).

Furthermore, the AccGSR and the peaks are significantly different depending on who the participants are interacting with. Moreover, how the temperature varies in three regions of interest across the face (i.e., forehead, left, and right periorbital regions) is a good indicators of how ready an individual is to react in an unforeseen situation.

7.7 Contribution

The main contribution of this work consists in the investigation of how individuals react in an unforeseen situation (i.e., while playing the Jenga game). Furthermore, to the best of our knowledge, the fight/flight system has not been used before in HRI. We have shown that there is a relationship between the FFFS system and the variation of the different physiological parameters.

This work was published in the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2018), “Oh! I am so sorry!: Understanding user physiological variation while spoiling a game task”, (Agrigoroaie et al., 2018a).

Chapter 8

Assistive applications (I)

8.1	Introduction	97
8.2	Related work	98
8.3	The ENRICHME project	99
	8.3.1 Robotic system	100
	8.3.2 Graphical User Interface	101
8.4	Lessons learnt from a first interaction with the elderly	102
	8.4.1 Scenario	103
	8.4.2 Interaction, data recording and analysis	103
	8.4.3 Thermal data	105
	8.4.4 Results and comments	106
	8.4.5 First lessons learnt	107
	8.4.6 Discussion	108
	8.4.7 Conclusion	108
8.5	Results of a 5-day interaction scenario designed for the elderly: a pilot study	108
	8.5.1 Scenario	108
	8.5.2 Participant description	110
	8.5.3 Physiological data extraction and analysis	110
	8.5.4 Hypotheses	110
	8.5.5 Results	110
	PANAS	111
	Digit Cancellation	111
	Integer Matrix Task	114
	Stroop game	115
	Hypothesis H1.	116
	Hypothesis H2.	117
	Other results	118
	8.5.6 Discussion	118
8.6	ENRICHME project testing phase	119
	8.6.1 RQ1: Which is the most used GUI application?	120
	8.6.2 RQ2: Which is the most played cognitive game?	120
	8.6.3 Performance analysis	120
	8.6.4 Discussion	123
8.7	Conclusion	124
8.8	Contribution	124

8.1 Introduction

The physiological internal state and cognitive games developed throughout the thesis were initially tested in laboratory settings. In order to see how these can be applied in

real-world environments, two vulnerable populations were selected to test our system. The two populations considered were: the elderly with cognitive impairments and individuals suffering from different sleep disorders. The elderly individuals who tested the system were part of the ENRICHME H2020 European project testing users, while the individuals suffering from sleep disorders are outpatients of a Sleep Unit from the University Hospitals Pitié Salpêtrière - Charles Foix in Paris. In this chapter, we are going to present the results from the study with the elderly individuals, while in Chapter 9, we are going to present the results with the individuals suffering from sleep disorders.

8.2 Related work

During recent years, development of innovative technologies for a sustainable and high quality home care of the elderly has become of primary importance. Extensive research has been carried out in the area of mobile autonomous robot companions operating in domestic environments.

The FP7 CompanionAble project (Merten et al., 2012) tried to address the issues of social inclusion and home care of persons suffering from Mild Cognitive Impairment (MCI) prevalent among the older population. The project focused on combining the strengths of a mobile robotic companion with the advantages of a stationary smart home. The FP7 Mobiserv project (Nani et al., 2010) developed an integrated and intelligent home environment for the provision of health, nutrition and well-being services to older adults. The goal has been to develop and use up-to-date technology like a companion robot, smart home, and smart clothes in an intelligent and easy to use way to support independent living of older persons. H2020 Mario project (Christos et al., 2017) aims to help people with dementia by enabling them to stay socially active using touch, verbal, and visual human-robot interaction tools. Another H2020 project called GrowMeUp (Martins et al., 2015) focuses on allowing the elderly people to live for longer in their own environment without losing contact with their social circles, staying socially active either via teleconference or through the social facilities provided within the system. Additionally, they aim to have behavior analysis components that will detect and report any emergencies and respond to the person's non-verbal cues and commands. The RAMCIP project (Kostavelis et al., 2015) aims to develop a service robot, capable to assist older persons in a wide range of daily activities, being at the same time an active promoter of the user's physical and mental health by deciding when and how to assist the user. It is accomplished either by initiating a multi-modal human-robot communication or by fulfilling a robotic manipulation task. The aim of the HOBBIT European project (Papoutsakis et al., 2013) was to use a robot in order to visually observe and interpret the actions of the elderly so that appropriate assistive services can be provided. The Robot-Era¹ (Di Nuovo et al., 2017) robotic system was designed as a mobile personal service robot with the aim of helping the elderly in their daily-life activities and to stimulate mental and physical exercise. The MoveCare² European research project developed a robotic system capable of providing with the help of artificial intelligence assistance and transparent monitoring for the elderly in their own homes.

¹<http://www.robot-era.eu>

²www.movecare-project.eu

8.3 The ENRICHME project

This work was part of the Horizon2020 ENRICHME European project³, whose purpose was to develop a socially assistive robot for elderly with mild cognitive impairment (MCI). The main aim of the robot was to help the elderly remain active and independent for longer and to enhance their quality of life. The project had partners from Italy, United Kingdom, France, Netherlands, Poland, Greece, and Spain. The system was tested in two Ambient Assisted Living (AAL) laboratories and in three elderly housing facilities.

The AAL laboratories are Fondazione Don Carlo Gnocchi, in Milano, Italy, and Stichting Smart Homes in the Netherlands. In both facilities 15 older people took part in the testing phase. The elderly housing facilities are Lace Housing Ltd, in the United Kingdom, Osrodek Pomocy Społecznej in Poland, and Aktios Licenced Elderly Care Units in Greece. In each of the three sites 4 elderly interacted with the ENRICHME robot.

The robot chosen for this project was the TIAGo robot developed by PAL Robotics (presented in Chapter 3.2). In each of the testing centers two robots were deployed. The system was simultaneously tested in all testing sites. The inclusion criteria for the elderly individuals to test the ENRICHME system were: be over 65 years old, be diagnosed of MCI by a neuropsychologist or be identified with mild cognitive dementia on the MMSE score (Folstein et al., 1975). Another aspect that was considered was the independence in executing basic activities for daily living assessed through Barthel ADL index (Mahoney et al., 1965). And lastly, the people needed to be living in one of the studied facilities.

A questionnaire was designed to collect opinions about robot-related requirements and to check its usefulness: Users' Needs, Requirements and Abilities Questionnaires (UNRAQ) (Cylkowska-Nowak et al., 2015). Data was collected in France, Greece, Italy, the Netherlands, Poland and the UK. Based on the results from the UNRAQ questionnaire and for the purpose of testing the developed system, multiple use cases have been developed. One of them is summarized below. This use case is related to games and entertainment provided by the robot to the user.

Jack likes to be active but due to his MCI, he finds it difficult to initiate activity. Sometimes, the robots' activity sensors and the sensors around the house detect that Jack is agitated and anxious. In order to reduce this anxiety state, the robot comes next to Jack, and suggests him to initiate a cognitive activity or to give a call to a friend that can reassure him. The system knows from past experiences that Jack does not like to be approached from the back, so it approaches him slowly and from the front. Jack likes to play Trivia game and ask the robot to play with him. The robot adapts the difficulty level as a function to his previous performance and the current emotional state. After some games, Jack can relax thanks to the robot that assisted him to engage in this activity.

Some of the most important needs identified are:

- for the robot to call somebody in case of an emergency
- to remind the elderly about the medication that needs to be taken
- to increase the safety in the home by monitoring the environment and suggesting improvements

³<http://www.enrichme.eu>



FIGURE 8.1: GUI main menu

- to increase the safety of the elderly by monitoring physiological health parameters like heart rate, body temperature, respiration rate
- to help the elderly preserve their memory function by playing memory games with them
- to help them find lost objects
- to remind them about certain appointments
- to encourage and guide in performing physical exercises
- to remind about meal times and drinks and to provide advice about a healthy diet
- to encourage the elderly to enhance their contacts with friends

8.3.1 Robotic system

Additionally to the sensors described in Chapter 3.2, the system also features an RFID reader module (ThingMagic Mercury6e) mounted below the touchscreen to detect objects with pre-attached RFID tags, as well as an environmental sensor (UiEM01) which provides temperature, humidity, light and air quality measurements. Smart Home Sensors (Fibaro Motion Sensor FGMS-001) for ambient temperature and humidity or electricity usage (Fibaro Wall Plug FGWP-102) were mounted in each room in the house. Data from these sensors were recorded on a server installed in the house. This server also served as a tunnel for a cloud-based application through which caregivers could interact with the robot and visualise the collected data.

For every-day interactions with the elderly user, a web-based graphical user interface (GUI) was developed for the touchscreen mounted on the torso of the robot. The GUI features nine custom developed applications: Cognitive games, Health tips, Physical activities, Find object, Environmental data (i.e., robot sensors; house sensors), Agenda, Call somebody, Weather forecast, and News reading. These applications (also shown in Figure 8.1) were designed based on the use cases of the ENRICHME project and were customized based on the feedback received from the early

test participants of the project. Each application is briefly presented in the following section.

8.3.2 Graphical User Interface

Cognitive games: as the end-users are elderly with MCI, we designed and developed nine cognitive games for helping them maintain and even improve their cognitive abilities. A review of the existing cognitive games designed for the elderly showed that none of them met the requirements of the current project. Two computer based solutions are represented by the M3W⁴ and the Sociable⁵ projects. Both of them contain games that address different cognitive categories: attention, language, memory, executive functions, logical reasoning, orientation. Most of the games are black-boxes (e.g., Sociable), therefore, do not provide access to the performance throughout the game, and only at the end of the game provide a final score. We want to be able to provide customized feedback at each step of each game; as well as a customized performance score at the end of each game. Other cognitive games platforms are not free to use (e.g., Luminosity) and are not easy to integrate in the current architecture of the system. Taking all these into consideration and based on the input from the experts and the early test participants, we have decided to implement nine cognitive games, each with multiple difficulty levels.

The games were designed to train memory, attention, and mental calculations, among others. The nine cognitive games are:

- *Digit and Letter Cancellation:* in a given time all occurrences of a randomly chosen digit/letter have to be found in a list of digits/letters.
- *Integer and Decimal Matrix Task:* the user needs to find in as many matrices as they can the pair of digits (integers or decimals) whose sum equals 10.
- *Memory Game:* the user first needs to memorize a set of images and then to decide if any of them are part of a new set of images.
- *Hangman:* this game is the implementation of the classical hangman game, where the user needs to guess a word by using letters of the alphabet.
- *Speed Game:* in this game two pairs of letter+number are shown on the screen. The user has to select as quickly as possible the pair that represents vowel+even number.
- *Puzzle:* the user has to reconstruct an image by dragging and snapping together puzzle pieces.
- *Stroop Game:* the user has to press on the button that corresponds to the color of the text that is displayed on the screen.

Health tips: In this activity, the user can listen to a series of healthy eating tips presented by the robot. The user can select to go to the next/previous tip or to listen to the same tip again. Currently the system provides nine tips related to a healthy diet: eating more fruits and vegetables, eating more fish, staying hydrated. The healthy tips are based on the British Nutrition Foundation recommendations⁶.

⁴<https://m3w-project.eu>

⁵<http://cognitivetraining.eu>

⁶<https://www.nutrition.org.uk/healthyliving/healthydiet.html>

Physical activities: For maintaining their physical abilities, the GUI features a physical activities application. The user is presented with a series of upper body exercises (i.e., head, shoulders, arms). Both auditory and visual feedback is provided by the robot for each exercise. The application uses the pose of the user as perceived by the RGB-D camera using the PrimeSense OpenNI body tracker. Based on the position of 22 joints, the angles between the joints of the upper body are computed before providing the feedback for each exercise. The physical exercises were decided together with the physical therapists from the care facilities.

Find object: The user can use the ENRICHME robot to localize different objects around the house. A visual interface was designed for the RFID-based object localization module of the robot. The user selects the object that he/she wishes to find. The robot will display the rooms/regions in descending order of the probability of finding the object in that room/region.

Environmental data: In this application, the user can select between visualizing the data from two types of sensors: an environmental monitoring sensor mounted on the robot and the Smart Home Sensors mounted around the house.

Agenda: The purpose of this application is for the user of the ENRICHME system to be able to visualize the appointments and reminders set up by the caregivers.

Call somebody: This application enables the user to perform video calls. The application was custom designed using WebRTC technology. The user has also the option of sending text messages to their contacts that are saved in the phone book of the application.

Weather application: In this application the user can see the weather forecast for its location. There are two visualization options: hourly forecast and daily forecast. The application uses the API of openweathermap.org.

News reading: The application uses the RSS 2.0 feed of different newspapers to gather the latest news and display them to the end user. The newspaper and the news category can be selected by the user. The news are presented both written and can be read out loud by the robot.

The ENRICHME system was deployed in the house of the elderly users. In a typical usage scenario, when the user wants to play some cognitive games, they approach the robot. If the robot is located in a different room, the user selects the Follow mode (i.e., a mode in which the robot follows the user) on the touchscreen interface and goes to a comfortable place to play the games (e.g., the living room sofa). From time to time the caregivers can also be present, but most of the time the user is alone when interacting with the ENRICHME robot.

8.4 Lessons learnt from a first interaction with the elderly

This first interaction with the elderly took place in LACE Housing elderly care facility in the UK during the first year of the ENRICHME project. This being the first interaction between the possible final users and the ENRICHME system, only two elderly individuals were recruited to interact with the robot and to test the modules that were created so far.

The robot used for this interaction was the Kompai robot designed by Robosoft⁷ (see Figure 8.2). The robot was equipped with an ASUS XTion Pro RGB-D sensor mounted beneath its head, an Optris PI450 thermal camera mounted on the head, and

⁷www.robosoft.com



FIGURE 8.2: The robot in one of the user's home

different sensors for measuring environmental parameters (e.g., ambient temperature). The torso of the robot features a touchscreen to facilitate user interaction.

8.4.1 Scenario

For this interaction, one of the use cases developed for the project has been implemented. In this use case, the activity sensors of the robot and the sensors located in the house detect that the person is restless, and the robot approaches the person to propose an activity for them to do. The possible activities are: playing a cognitive game, doing some physical exercises, or learning some healthy eating tips.

For this scenario, the robot uses a leg detection algorithm by using the laser range finder mounted on the robot (Dondrup et al., 2015) so as to track people in the room. Based on data from an online survey (Ferland et al., 2017), where participants would define approach parameters (i.e., stopping distance, deceleration, and curvature of the path) according to the robot's personality, the robot was configured to approach the person in a submissive and friendly manner. The robot automatically approaches the closest person it detects, and keeps itself oriented toward that person at a maximum distance of 1.0 m. Therefore, even if the person is moving in the room, the robot is able to continuously follow the person.

8.4.2 Interaction, data recording and analysis

A face recognition module based on (Amos et al., 2016) was trained with the images of the two residents. Therefore, the robot was able to recognize them and greet them by their name when they were in range. Next, an early version of the web-based graphical user interface (see Figure 8.3) was displayed on the touch screen of the robot.

The cognitive games available on the robot at that time were: Digit cancellation, Puzzle, Hangman, and the Memory game. Feedback was given after each game, summarizing the user's performance based on the hints used, errors made, total game time.

The physical exercises activity consisted of three parts, each with a duration of 1 minute. In the first part, the robot used only speech to tell the user what exercise to



FIGURE 8.3: The developed interface with the potential activities

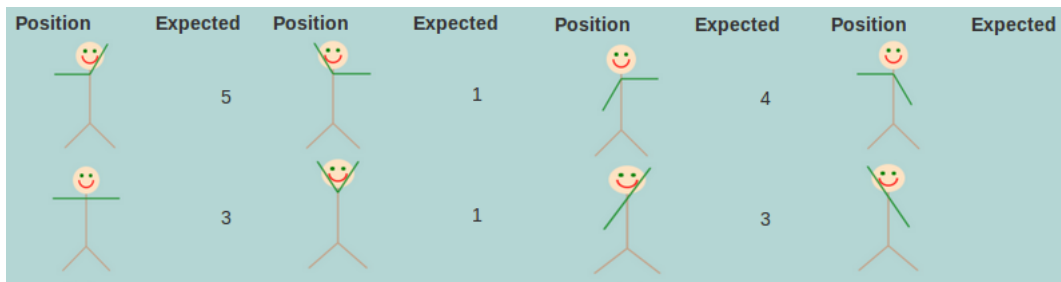


FIGURE 8.4: The physical exercise feedback

do. In the second part, the user was only shown images of the exercises they have to do. While, in the third part, the robot used both speech and images for the exercises. The exercises were randomly chosen from a set of 8 possible arm exercises shown in Figure 8.4. These exercises were selected in collaboration with the physical therapist available in the care facility. We assured ourselves that the participants were able to perform these physical exercises. After the execution of the physical exercises, the user was informed about the number of times they correctly performed each exercise.

The third proposed activity was the health information tips. The robot presented the user 8 healthy eating information such as not skipping breakfast or eating less refined sugars. At the end, the amount of information retained was tested with a short quiz.

After the interaction, each participant was given two questionnaires: the BIG5 personality questionnaire (Goldberg, 1990) and a custom made questionnaire developed for finding out what the participants liked or disliked about the graphical interface and the overall robot behavior. Each questionnaire took between 5 and 10 minutes to complete and consisted of yes/no questions or questions with a 5-point Likert scale (1 - “strongly disagree”, 5 - “strongly agree”).

Throughout the interaction between each participant and the robot, the data from multiple sensors was recorded. For human robot interaction, the data sources used are the RGB-D data, the thermal image, the audio, and the data from the skeleton tracker. For the audio an external omnidirectional microphone was used, and the skeleton tracker used the OpenNI NiTE library with the ASUS XTion Pro sensor. The data from the skeleton tracker was used for determining if the elderly executed correctly the physical exercises. Except for audio, the Robot Operating System (ROS) framework (Quigley et al., 2009) was used for robot behavior coordination, data recording and analysis.

For compactness, the following notation is used: P_i for participant i (1 or 2), E_j for event j (1 to n), and P_iE_j for the j th event that was recorded for Participant

i. The possible events are: performing the physical exercises, playing one of the cognitive games.

Participant P1 is a 73 years old introverted male who suffers from Parkinson's disease and has cerebellar ataxia (the inability to coordinate balance, gait, extremity, and eye movements). Participant P2 is a 83 years old extroverted female and has arthritis. Both participants started with the health tips, continued with the cognitive games, and ended with the physical exercises. The total recording times for P1 and P2 are 37 and 29 minutes, respectively. Out of the total interaction time, two events for each participant were chosen for further analysis. These events were chosen because both participants showed a strong reaction towards something they disliked or liked. Before the interaction started with the two participants, there was a short demo for all the residents of the care facility. One of the games that was tried during the demo was the puzzle game, which was found interesting, but difficult by most of the residents. The level of the game shown in the demo was the medium one, with 30 puzzle pieces. Just by seeing them they thought that they would not be able to complete the puzzle. When participant P1 saw the puzzle game among the possible cognitive games, there was a strong reaction: he found it too difficult and did not want to try it. Data during this reaction was chosen to be P1E1. P1E2 is represented by the last two minutes from the physical exercises activity. The first minute of the physical exercise was not chosen due to a lot of movement of participant P1. There was nothing shown on the screen during this time, therefore he turned towards the other people in the room.

Participant P2 firmly expressed a strong dislike towards quizzes, therefore at the announcement of having to complete a quiz after the healthy eating tips a strong reaction was observed. Therefore, this was considered as P2E1. P2E2 consisted of the participant playing the puzzle game. As P1E1 consisted of a strong reaction on a game, for P2 was extracted an event related to a game in order to see if the temperature variation is similar for both participants P1 and P2, respectively.

We wanted to investigate if facial temperature variations and speech variations could be observed while the participants reacted to these events, and thus could be recognized by the robot in the future.

Participant P1 wore glasses throughout the interaction, during the memory game he changed his original glasses with another pair to better see the text on the screen. Participant P2 started the interaction without glasses, but during the puzzle game glasses were put on. During both interactions with the robot, other people were present in the room, and sometimes interacted with the participants. This resulted in situations such as a loss of facial temperature data because the participant turned his/her head, or voices other than the participants' being recorded by the external microphone. These events were not discouraged by the experimenters as it was believed that they would provide realistic information on situations that were certainly going to re-occur in future tests.

8.4.3 Thermal data

Facial temperature was extracted from 6 regions of interest (see Figure 8.5): the forehead, the region around the eyes, the nose, the tip of the nose, the perinasal region, and the mouth. In post-analysis, the whole face was selected manually at the beginning of each event, and a correlation tracker based on facial landmarks implemented with Dlib (King, 2009) was used to track the regions of interest. This represents one of the earlier versions of the facial temperature extraction methodology presented in Chapter 4.2.

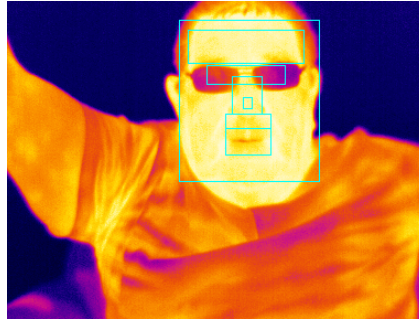


FIGURE 8.5: The regions of interest shown for the first participant. Temperatures range from 20°C in dark purple to 40°C in light yellow.

When participants turned their heads to speak with other people present in the room, sudden drops in facial temperature would appear. These were caused by background pixels entering briefly the tracked ROIs. Furthermore, as it can be seen from Figure 8.5 in the eyes region when the participant is wearing glasses, the temperature is very low. Therefore, readings associated with the glasses or the background have to be ignored. To do this, a histogram of the temperatures was plotted.

It was observed that the histogram was trimodal, with a minimum between the last two modals at 33°C . The first modal represents the background temperature (with a mean of 26°C), while the second one represents the glasses (with a mean of 30°C). Therefore all temperatures lower than 33°C were discarded. The same consideration was made for the nose region, as part of that region also contains the glasses. For the other regions the same processing was applied. The histogram was bimodal with the minimum between the two modal at around 30°C . The first modal (with a mean of 27°C) represented the background, therefore all values below 30°C were discarded. The average temperature of each region for each frame was computed and then filtered using a low pass Butterworth filter. The sampling frequency for the data is 9Hz, and a cutoff frequency of 0.64 Hz was chosen ((Yu et al., 1999), equation(9)) so as to filter out the variations in the temperature that are due to the small movements of the head or of the camera. For a better understanding of the general trend of the temperature over time a linear least-squares regression was applied (e.g., the eyes region for P1: slope of 0.00034, regression intercept of 35.27 and correlation coefficient of 0.15).

8.4.4 Results and comments

When looking at the raw thermal data we can observe several things from the very beginning. First of all, it can be easily observed that the presence of glasses produces a drop of temperature in the orbital region. However, this can be used to detect glasses and adapt the behavior of the robot accordingly. For example, if the user wants to play a cognitive game and the robot detects that the user has no glasses it can suggest wearing them based on previous information that the user sees and performs better with them.

When having a closer look at the mouth temperature variation in correlation with the audio, and the RGB-D data, it can be detected when a user has spoken or kept their mouth open. For example, in Figure 8.6 all the variations are due to the participant speaking. Even if all temperatures that correspond to the background have been eliminated, at timestamp 00:28 there is still a small drop due to the turning of the person. This can appear due to measurement errors as the tracker does not follow properly the facial features. The region of interest can include other parts of

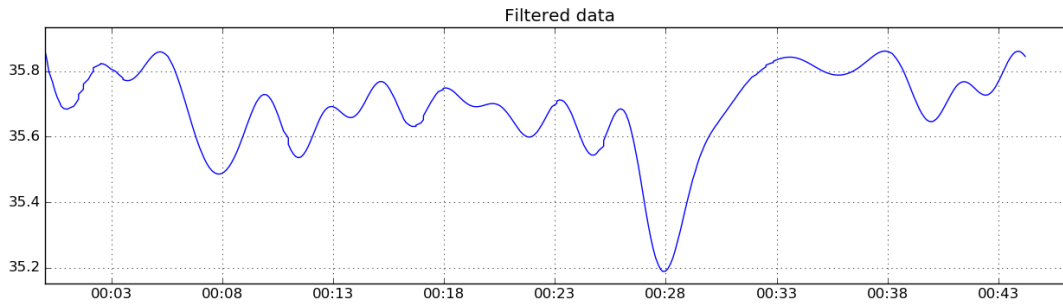


FIGURE 8.6: Mouth temperature variation

the face (e.g. for the mouth it can also include parts of the cheeks or the chin). The thermal data can be used individually, but in correlation with the RGB-D and audio data it can provide more robust information to the robot.

This information could help the robot better adapt its behavior in different situations (e.g., not to interrupt if the owner is speaking).

When comparing the two events in which both participants, P1 and P2, had a visible reaction to an unwanted event the following were observed. The duration of the event for participant P1 was of 25 seconds, while for participant P2 it was of 39 seconds. For participant P1, the temperature in all 6 regions increased between 0.05°C for the eyes region and up to 0.6°C for the tip of the nose. For participant P2, the temperature for the nose and the tip of the nose remained constant over time and it decreased for the other regions; between 0.01°C for the eyes and up to 0.5°C for the mouth region. According to (Ioannou et al., 2014b) the temperature variation for participant P2 are indicators of mild stress, while for participant P1 the increase of temperature in the forehead region is an indicator of stress. These are some preliminary results. More tests will take place in the near future with more participants and for longer periods of time.

8.4.5 First lessons learnt

Some of the general lessons that were learnt from the interaction can be summarized as follows:

- The position and the orientation of both RGB-D and thermal camera is very important as in some situations it can miss important information. The cameras should be able to cover the face of the person regardless if it is standing or sitting.
- There will be situations in which the user will not face the camera directly. In those situations the data should be discarded.
- Users might have glasses; the approach for extracting the periorbital temperature should take this into consideration as this region provides paramount information about the user.
- For a better detection of the facial landmarks, the RGB image should be used. With proper inter-camera calibration, coordinates of the facial features can be transposed from the RGB-D camera to the thermal camera. Thus, a smaller region of interest can be defined so other parts of the face are not included in case of sudden movements.

8.4.6 Discussion

Based on the questionnaire regarding their likes and dislikes, the two participants enjoyed playing the games (approval of 4.2/5) except the puzzle game (approval of 1/5). They liked the layout, and the size of the buttons was large enough for them. Remarks were made for the Puzzle game. The participants and other residents found that 30 pieces of puzzle for medium level were too many. Therefore the number of pieces needs to be adjusted, 30 pieces of puzzle will be used for the hard level, as suggested by the participants. Based on the questionnaire regarding their thoughts on the overall behavior of the robot, both participants found the robot's approach behavior natural (approval of 5/5), in its approach it was neither dominant nor submissive (approval of 3/5), and appeared to be friendly (approval of 4.5/5). Participant P1 found the robot a little extroverted, while participant P2 found it neutral.

8.4.7 Conclusion

In this section, we have presented some first lessons learned from a first interaction between the ENRICHME robot and two residents in a care facility (Lace Housing Ltd in the United Kingdom). The robot interacted with the participants in their own homes. Events like turning their heads and talking to other people in the room were not discouraged by the experimenters. These events provide useful information on what kind of situations might occur in the long-term testing of the system. From the results of this work, we can posit that the facial temperature variations can be observed. The robot could adapt its behavior based on these variations to provide a more natural interaction with the elderly. For example if during a game the robot detects a high stress level (Sorostinean et al., 2015) it could adjust the difficulty level of the game.

Based on the information gathered in this first interaction, the graphical interface and the applications were improved. Furthermore, new applications were developed and included in the ENRICHME system. In the following sections we are going to present some of the results of the interaction between the final end-users of the ENRICHME project and the robotic system developed.

8.5 Results of a 5-day interaction scenario designed for the elderly: a pilot study

8.5.1 Scenario

For this experiment, two elderly participants were recruited from an elderly housing facility in the United Kingdom (UK). Both participants tested the ENRICHME system for a duration of 10 weeks. The experiment took place in their own houses, which are part of the housing facility. For the experimental procedure a 3 days interaction were planned with each participant. However, due to the degradation of the visual capabilities of one of the participants, the experiment could be carried out with only one participant. In the end, there were 5 days of interaction with only one participant.

Each day of interaction consisted of two sessions: one in the morning (around 11 am) and one in the afternoon (around 3 PM). The tasks performed during each session were the same. Each interaction took place in the living room of the participant (see Figure 8.7a).

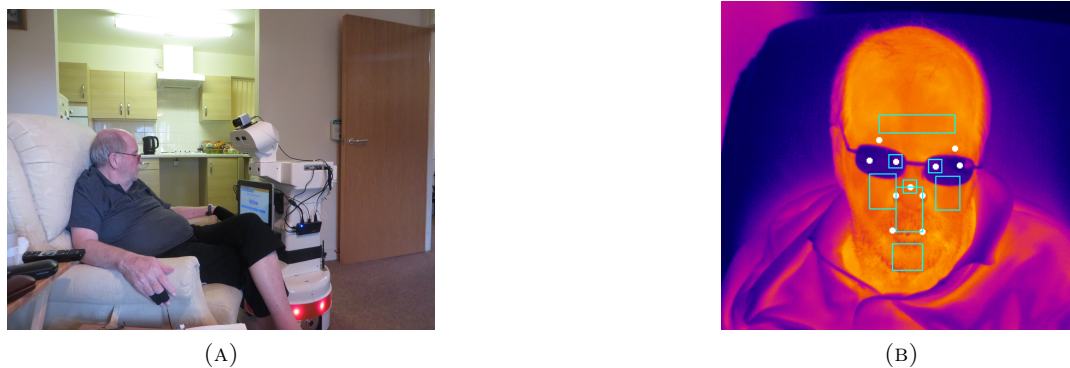


FIGURE 8.7: (a) Experimental setup; (b) Facial feature points and ROIs

As the participant was not familiar with the experimenter, two days of visits were planned. In the first day, the experimental setup was put in place. While on the second day, the two interaction sessions took place. The data on this day was not considered for the analysis as we wanted to eliminate all influences of different novelty parameters (e.g., presence of new experimenter, the experiment, the difficult levels of the cognitive games). During the first day, the participant also filled out multiple questionnaires: EPQ, MEQ, RST-PQ and the PANAS questionnaires (see Chapter 4.5).

Each interaction session consisted of three main phases: Relaxation, Cognitive Games, and News Reading.

- **Relaxation:** each interaction session started with a relaxation period of three minutes. On the first day, the participant was asked to choose one song that he would like to listen to that had a relaxing effect. For the following sessions the same song was played by the robot. The purpose of this phase was to relax the participant before the following tasks.
- **Cognitive games:** the participant had to play 5 cognitive games. The order of the games was chosen by the participant, and each game had to be played only once. The cognitive games are: Digit Cancellation (two difficulty levels), Integer Matrix Task (two difficulty levels), and Stroop game.
- **News Reading:** as a continuation of our previous study (Agrigoroaie et al., 2017c), we tested the news reading application with two interaction distances (i.e., 70cm, and 100cm) and different sensory stimuli (either only visual (C1), or a combination of visual and auditory stimuli (C2)). The conditions and interaction distance were chosen in a random order. Each interaction session consisted of a total of eight news.

Before the relaxation phase and after the news reading phase, the participant filled the PANAS questionnaire. After each interaction session, the participant had to also fill in the custom developed post-questionnaire.

During all tasks the physiological data was also recorded. The RGB-D and thermal cameras were already mounted on the robot, while the GSR sensor was placed on the right hand of the participant (even if the participant is right-handed, due to a shoulder injury, the left hand was used for the interaction).

For this study, we are interested both in the performance during the cognitive tasks, as well as in the variation of the physiological parameters during all tasks.

8.5.2 Participant description

The participant is a 69 years old male from the United Kingdom. He is a retired self employed decorator. He used to live alone, but he moved to the housing facility after he fell and he broke his hip. Not very long ago, he injured his right shoulder, and even though he is right handed he had to use his left hand for interacting with the touchscreen of the robot. He has no previous knowledge about robotics, except the news that he sees on the television. He has been smoking for more than 40 years and he likes to drink coffee. He does not drink any alcohol.

The results from the questionnaires show that he is very extraverted (a score of 12 out of 12), and that he has a low neuroticism level (a score of 2 out of 12), he is an intermediate type when considering the morningness-eveningness questionnaire, with a score of 56. The MEQ score can be between 16 and 86. Evening types have scores between 16 and 41; morning types have scores between 59 and 86.

His results from the RST-PQ show that his FFFS score is above average (3.3 out of 4), while his behavior inhibition score is very low (1.1 out of 4). Regarding the behavior activation system, for 2 of the 4 subscales (reward interest, and impulsivity) he scored below average (2.1 and 2.3 out of 4), while for the other two (reward reactivity and goal driven potential) he scored above average (2.6 and 3.1 out of 4).

8.5.3 Physiological data extraction and analysis

As seen in Chapter 2 the physiological parameters that are good indicators of an individual's internal state are: facial temperature variation, and GSR among others. For this research, we are going to extract the facial temperature variation from the thermal data, the accumulative GSR (AccGSR) and the number of peaks from the GSR data. The data extraction and analysis algorithms are presented in more detail in Chapter 4.

Heart rate and Respiration rate The HR and RR were extracted from the RGB data. For this purpose the algorithm presented in (Rahman, 2016) was used. First, the face of the participant is detected, together with 68 facial feature points. The forehead region is defined and the mean value of the green channel is extracted. The HR and RR can be extracted due to the periodicity of the data. For the analysis, we extracted the average HR (AvgHR) and average RR (AvgRR) for each of the tasks executed.

8.5.4 Hypotheses

For the variation of the physiological parameters we developed two hypotheses:

H1. The variation of the physiological parameters is dependent on the task performed.

H2. The difficult levels of the games are characterized by higher values of the GSR parameters.

8.5.5 Results

In this section, we present the results corresponding to the performance during each of the cognitive games, the results of the PANAS questionnaire, as well as the results for our hypotheses. As the participant could choose the order of the games, we noticed that the participant preferred to start with the easy levels of each game, then continue with the Stroop game, and at the end to do all difficult levels of the games.

PANAS

The PANAS questionnaire was filled in before the relaxation part and after the news part. For each session we computed the difference between the mood after the interaction and before the interaction. We wanted to see if there are differences between the morning and evening sessions, and between the interaction days.

For the positive affect difference, an one-way ANOVA analysis shows that the participant felt better after the morning session than after the evening session ($F(1,6)=0.94$, $p=0.368$). On days day2 and day3 the participant felt better after the interaction than in the last two days (day4, day5) ($F(3,4)=5.43$, $p=0.067$). This result can be related to the participant's health. On the last two days he had some health problems and was not feeling good at all.

No differences were found for the negative affect. The negative affect recorded was at a constant value of 10 for 7 out of the entire 8 interaction sessions considered for the analysis. The only day in which the negative affect increased after the experiment, was on day3 in the

evening session. As seen previously, an increased value for negative affect is associated with unpleasurable engagement. During that session, for the news reading application multiple technical difficulties were encountered, and the news phase had to be restarted multiple times.

Digit Cancellation

For the Digit Cancellation game three parameters were extracted and used for the analysis: total game time, average time to select an answer, and the total number of mistakes. The two factors used for analysis are the day when the sessions took place (day2, day3, day4, or day5), and the time of the session (morning, or evening). Figure 8.8a shows the variation of the total game time in each day for the easy level, while Figure 8.8b shows for the difficult level.

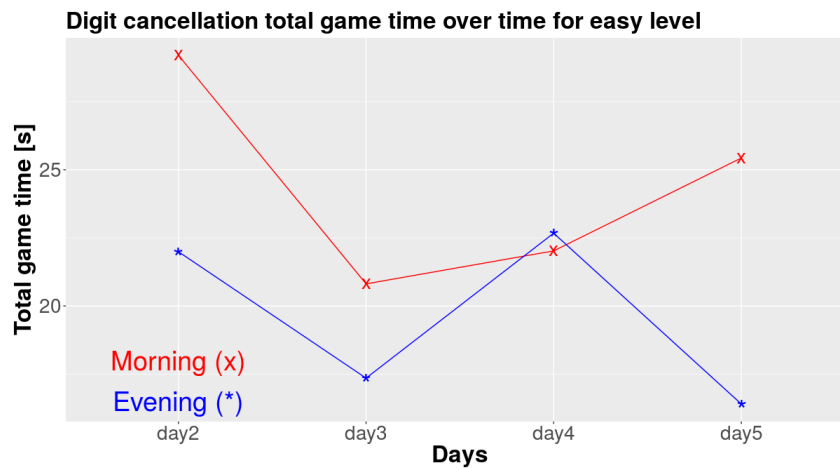
For the total game time, we found that in general our participant finished the easy level faster in the evening session than in the morning one ($F(1, 6)=3.72$, $p=0.1$) (Figure 8.9a). While for the difficult level the opposite was found ($F(1, 6)=1.73$, $p=0.23$) (Figure 8.9b). This is also due to the fact that the participant did not finish the difficult level 4 times out of the total 8 times taken into consideration for the analysis. No important differences were found for the total game time between the days.

For the average time to select an answer, we found that for the easy level in the morning hours the participant took longer to make a selection than in the afternoon hours ($F(1, 6)=0.96$, $p=0.36$). For the difficult level, the average time to make a selection was similar in the morning and evening sessions.

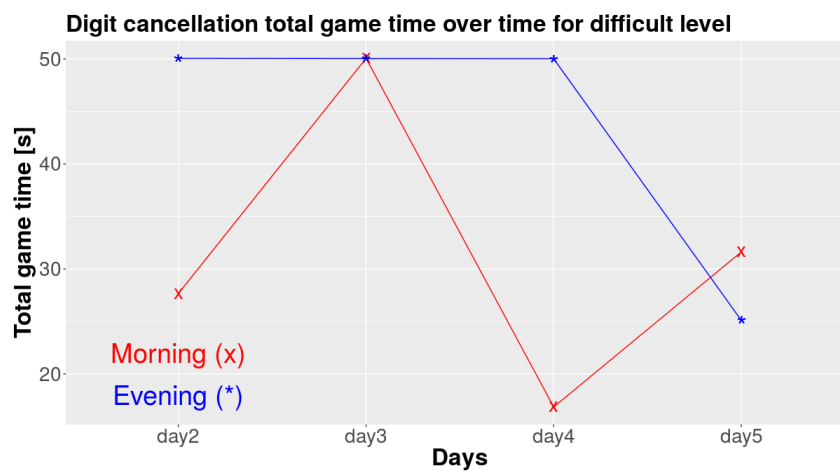
For the analysis based on the days, we found that for the difficult level, the participant took longer to make a selection in the first two days than in the last two days ($F(3,4)=3.32$, $p=0.13$) (Figure 8.10). No important results were found for the easy level.

During the easy level the participant made no mistakes. For the difficult level, the participant made the most mistakes in the evening hours ($F(1,6)=2.359$, $p=0.175$). Moreover, in the first two days more mistakes were made than in the last two days ($F(1,6)=2.359$, $p=0.175$).

Overall, even if no significant results were found, a tendency can clearly be seen. More interaction sessions and more participants could help find the significant results.



(A)



(B)

FIGURE 8.8: Digit Cancellation total game time over time: (a) Easy level; (b) Difficult level

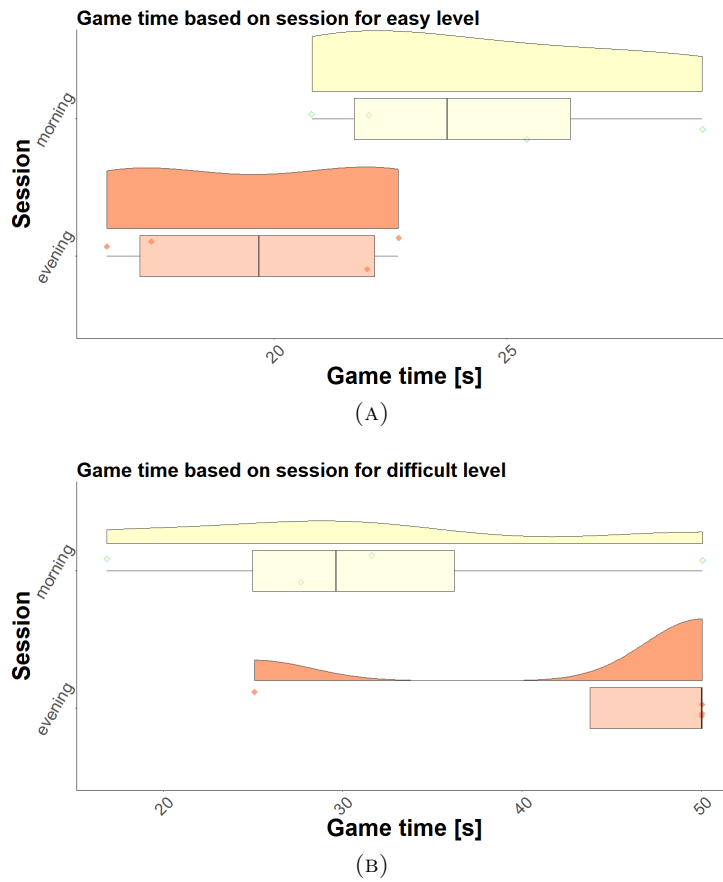


FIGURE 8.9: Digit Cancellation: (a) Easy level; (b) Difficult level

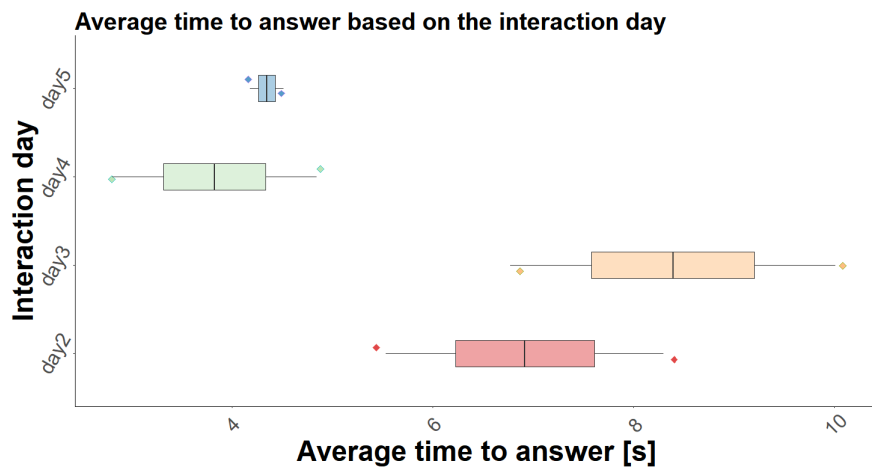


FIGURE 8.10: Digit cancellation average time to make a selection based on the day

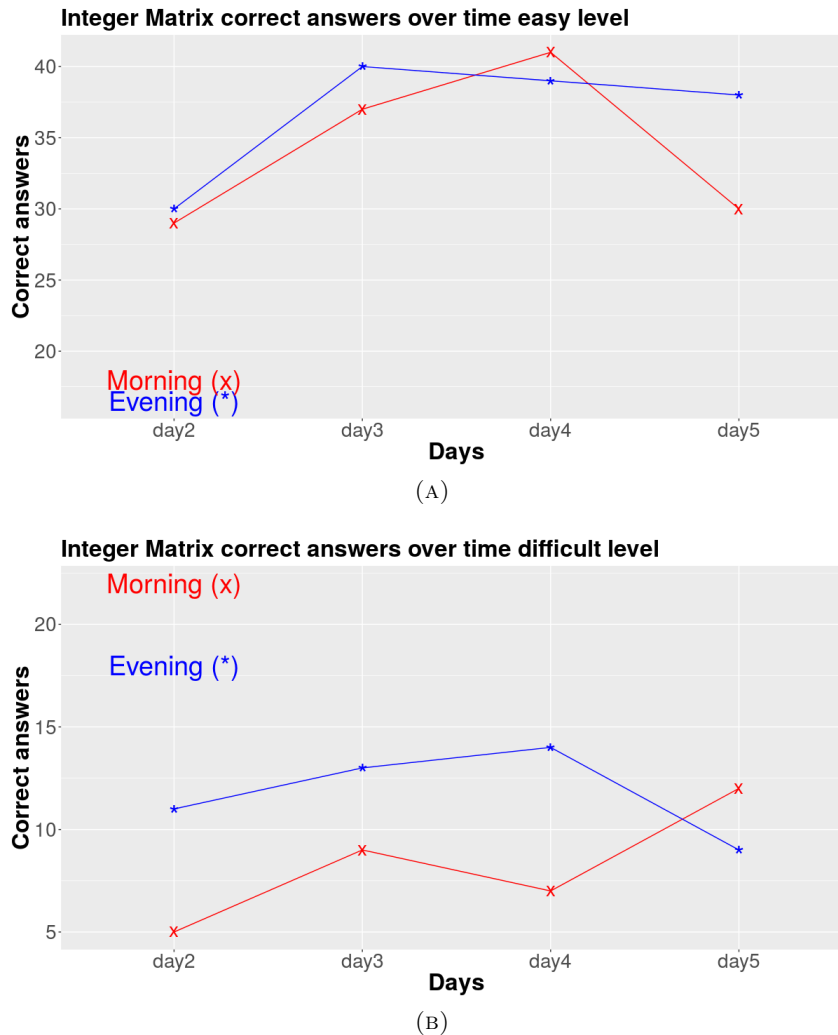


FIGURE 8.11: Integer Matrix correct answers: (a) Easy level; (b) Difficult level

The results of the post questionnaire show that the participant did not find the easy level neither difficult, nor stressing. While for the difficult level, the participant reported that he found the game difficult, but not stressing.

Integer Matrix Task

As the participant really tried not to make any mistakes, only one mistake was made in all 8 interaction sessions. Therefore, only the number of correctly solved matrices and the time to find the first of the two digits were considered for the analysis. Figure 8.11a shows the number of correctly solved matrices for each day for the easy level, while Figure 8.8b shows for the difficult level.

For the total number of correctly solved matrices no differences were found for the easy level between the morning and evening sessions. For the difficult level, we found that the participant solved more matrices in the evening than in the morning ($F(1,6)=3.54$, $p=0.1$) (see Figure 8.12).

We found that in the first day the participant performed worse in the easy level than in the other days ($F(3,4)=4.615$, $p=0.086$), especially compared to day3 ($p=0.22$) and day4 ($p=0.17$). No important differences were observed for the difficult level.

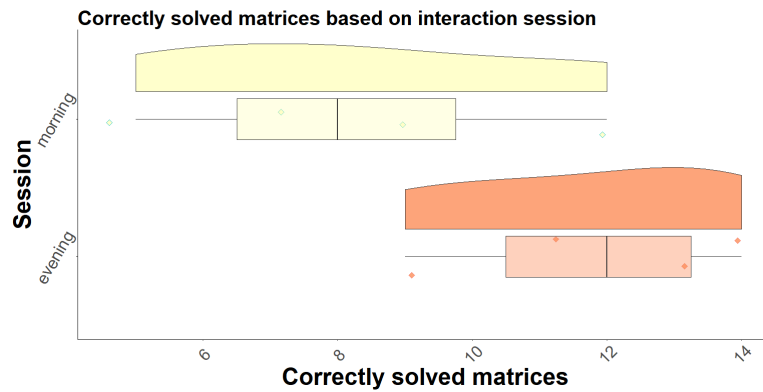


FIGURE 8.12: Matrix task correctly solved matrices in the difficult level based on interaction time

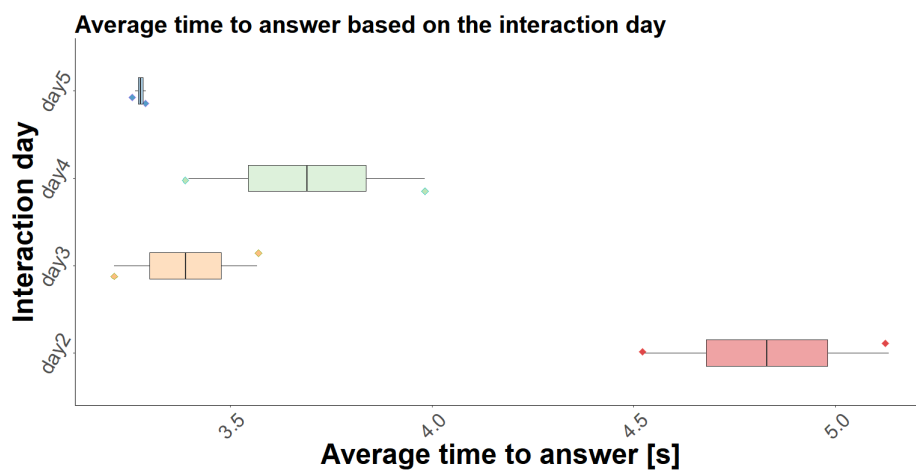


FIGURE 8.13: Stroop Game average reaction time based on interaction day

When considering the time taken to find the first answer the following results were found. For the difficult level we found that the participant found more quickly the answers in the evening session than in the morning session ($F(1,6)=2.315$, $p=0.17$). No differences were found for the easy level.

For the analysis based on the day of interaction, for the easy level, it took longer to find the first element in day2 than in any other day ($F(3,4)=5.4$, $p=0.06$). For the difficult level, no differences were found between the days.

The participant did not find the easy level difficult, nor stressing, while the difficult level was considered difficult but not stressing.

Stroop game

For this game we only considered for the analysis the average reaction time and the total game time. Only three mistakes were found throughout the 8 sessions.

No differences were found between the morning and the afternoon sessions for the game time and neither for the average reaction time.

On day2 it took longer, on average, to find the answer than on the other days ($F(3,4)=9.69$, $p=0.02$), especially when compared to day3 ($p=0.056$), and day5 ($p=0.052$).

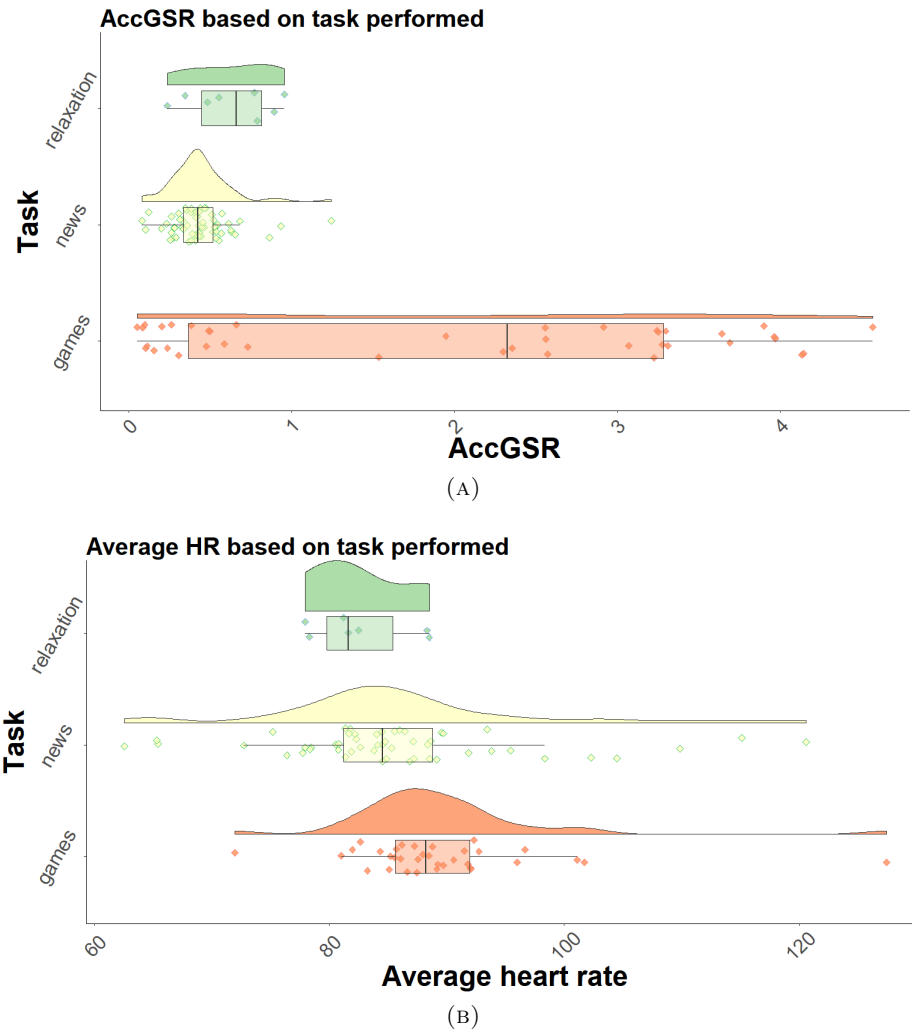


FIGURE 8.14: H1 results: (a) Boxplot for AccGSR; (b) Boxplot for AvgHR

Hypothesis H1.

For this hypothesis we state that *the variation of the physiological parameters is dependent on the task performed*. To test it, first, a normality test (Shapiro Wilks test) was applied on the data. None of the parameters are normally distributed, therefore a Kruskal Wallis analysis of variance was applied.

Significant differences were found for the number of GSR peaks ($\chi^2 = 15.97$, $p < .001$). A pairwise Wilcoxon rank sum test with a Bonferroni correction revealed significant differences between the games and relaxation ($p = .01$), and news activity and relaxation ($p < .001$). Another significant result was found for the AccGSR ($\chi^2 = 16.25$, $p < .001$) (see Figure 8.14a). A pairwise Wilcoxon rank sum test with a Bonferroni correction revealed significant differences between the news activity and games ($p < .001$).

A significant result was found for the AvgHR ($\chi^2 = 10.66$, $p = .004$). A pairwise Wilcoxon rank sum test with a Bonferroni correction revealed significant differences between the news activity and games ($p = .019$), and games and relaxation ($p = .023$). A boxplot of the results is shown in Figure 8.14b. In (Turner et al., 1985) the authors have found that the HR increases in a mental arithmetic task compared to a baseline level. Our results are thus in accordance with the literature, with higher HR levels

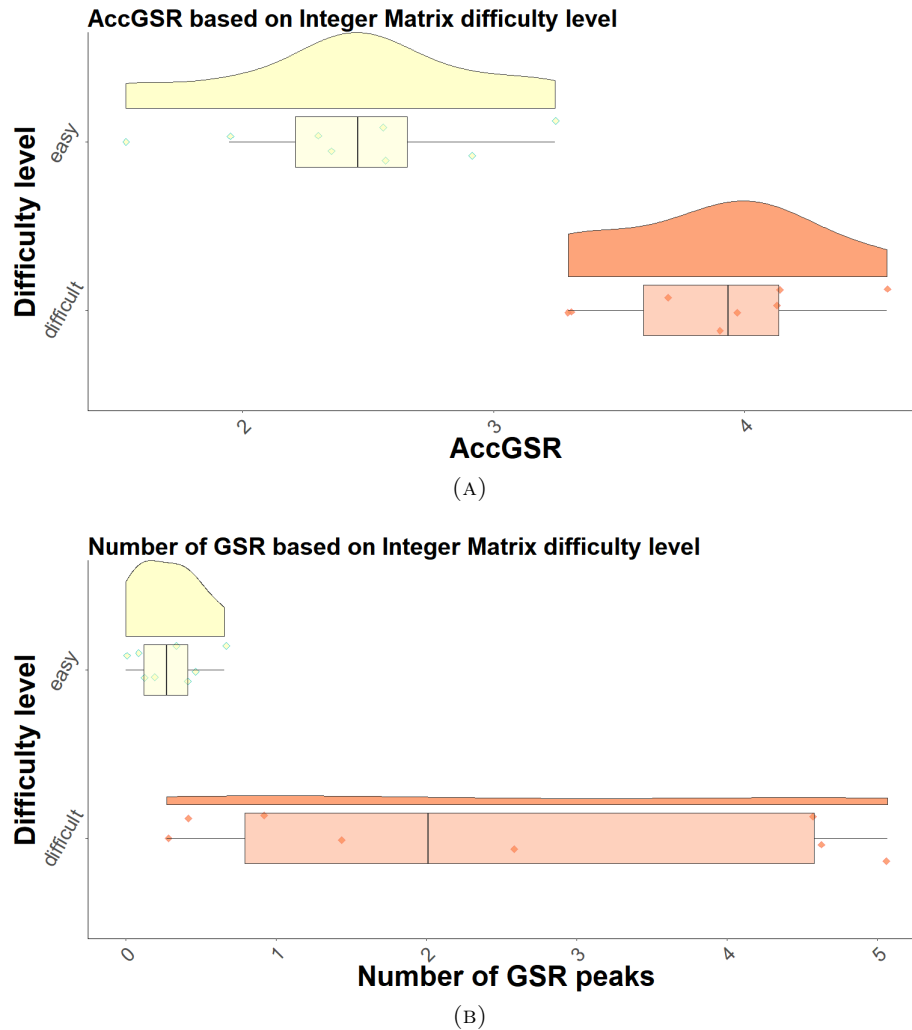


FIGURE 8.15: Integer Matrix Game: (a) AccGSR; (b) Number of GSR peaks

during the cognitive games compared to the relaxation phase.

Considering these results, we can state that **hypothesis H1 is confirmed**. By looking at the GSR parameters and the AvgHR we can differentiate between the tasks performed.

Hypothesis H2.

For this hypothesis we state that *the difficult levels of the games are characterized by higher values of the GSR parameters*. In order to test this hypothesis, we have to first select only the games that have multiple difficulty levels (i.e., Digit Cancellation and Integer Matrix Task).

For Digit Cancellation, we found significant differences for the AccGSR with higher AccGSR during the difficult level than the easy level ($F_{1,14} = 11.22, p = .004$). For the Integer Matrix Task, we found significant results both for the AccGSR ($F_{1,14} = 35.5, p < .001$) and for the number of peaks ($F_{1,14} = 9.47, p = .008$). Both parameters show higher values during the difficult level (Figure 8.15).

Based on these results, we can state that during the difficult levels of the games, significantly higher values of the GSR parameters were found. Therefore **hypothesis 2 is confirmed**.

Other results

We are also interested in investigating the correlation that exists between the games performance and the variation of the physiological parameters.

During the difficult level of Digit Cancellation, two positive correlations were found between the temperature variation in the nose region and the average time to find an occurrence ($p = 0.03, r(3) = 0.89$), and the total game time ($p = 0.01, r(3) = 0.93$). For the difficult level of the Integer Matrix Task the following significant results were found. A negative correlation was found between the number of peaks and the average time needed to solve a matrix ($p = 0.03, r(6) = -0.73$). In contrast, a positive correlation was found between the temperature variation in the forehead region and the average time needed to solve a matrix ($p = 0.01, r(6) = 0.8$). For the total number of matrices solved, there is a positive correlation with the number of peaks ($p = 0.02, r(6) = .78$), and a negative one with the temperature variation in the forehead region ($p = 0.01, r(6) = -0.83$).

8.5.6 Discussion

For this study, we only had one participant, therefore, the results that we obtained will have to be confirmed with more participants, including also females. This preliminary data shows that there is a tendency for this participant to perform differently in the morning and in the afternoon hours. Furthermore, we found evidence that there is an improvement over time.

The results from the Digit Cancellation show that the participant finished the game faster in the evening than in the morning for the easy level. No mistakes were made for the easy level, but some mistakes were made for the difficult one. Furthermore, for the difficult level, the participant did not finish the game 4 times out of the total 8 sessions. The number of mistakes decreased from the first two days to the last two days. We believe that with more sessions the participant would improve its performance even more.

Regarding the Integer Matrix Task, for the difficult level the participant solved more matrices in the evening than in the morning, and it took a shorter time to find the answer in the evening than in the morning. For the easy level, the participant showed an improvement after the first day, while for the difficult level, more sessions are needed to see a clear improvement.

A clear improvement over time was found for the Stroop game. After only one day the participant performed significantly better in the following days.

Our extraverted participant was more efficient at around 3pm than at around 11am, a result, which is in accordance with the theory proposed by Colquhoun (Colquhoun, 1960). For the morningness-eveningness level, we found that only for the Digit Cancellation difficult level the participant performed better in the morning than in the evening. Further studies have to be devised in order to investigate if the same results are found for other elderly with MCI.

One of the questions in the custom developed post-questionnaire asked the participant if he would have preferred to do the same tasks on a tablet. After each interaction he said that he preferred to interact with the robot.

As seen in the Results section, all our hypotheses are confirmed. The variation of some physiological parameters depend on the task being performed (AccGSR, peaks, AvgHR). The GSR parameters (i.e., AccGSR and number of peaks) can be used to differentiate between two difficulty levels of the games used in this study (i.e., Digit Cancellation and Integer Matrix Task).

TABLE 8.1: Testing duration for each participant

Participant	Site	Start date	End date	Total duration
P1	AKTIOS	11/10/2017	20/12/2017	10 weeks
P2	AKTIOS	21/12/2017	28/02/2018	10 weeks
P3	AKTIOS	11/10/2017	20/12/2017	10 weeks
P4	AKTIOS	21/12/2017	28/02/2018	10 weeks
P5	PUMS	02/11/2017	04/01/2018	9 weeks
P6	PUMS	08/01/2018	13/03/2018	10 weeks
P7	PUMS	14/10/2017	24/12/2017	10 weeks
P8	PUMS	02/01/2018	09/03/2018	9 weeks
P9	LACE	06/12/2017	08/02/2018	9 weeks
P10	LACE	08/11/2017	27/12/2017	7 weeks
P11	LACE	29/12/2017	16/02/2018	7 weeks

The correlation between the physiological parameters and the cognitive games performance parameters was also analyzed. The results show that there is a correlation between the two. Significant correlations between the performance and the variation of the physiological parameters were found only for the difficult levels. We can conclude that the difficult levels of the games were more cognitively arousing for our participant than the easy levels.

Our results show that in a human-robot interaction scenario it is important to also look at the physiological parameters of the individual that the robot interacts with. They could provide valuable information regarding the internal state of the individual. One limitation of the current study, is that it only looks at the physiological parameters variation for one elderly participant. A study with more participants is needed in order to confirm and validate our results. However, our study show a potential trend that needs to be more investigated. The participant stated that given the option to perform the same activities by using just a tablet, he would still prefer to interact with the robot.

8.6 ENRICHME project testing phase

In this section we present the results of the testing of the ENRICHME system with the final end-users for the project. The developed ENRICHME system was tested in the elderly case housing facilities (i.e., LACE, PUMS, and AKTIOS). On each site two robotic systems were deployed.

A total of 11 participants (4 males and 7 females) tested the ENRICHME robot. The testing duration for each robot ranged between 7 and 10 weeks (see Table 8.1). The ages of the participants ranged between 70 years old and 90 years old. Figure 8.16 shows three participants interacting with the ENRICHME system in each of the sites.

In this paper, we focus on the following research questions:

- RQ1: Which is the most used GUI application?
- RQ2: Which is the most played cognitive game?

Furthermore, we also investigated the performance over time for one of the participants.

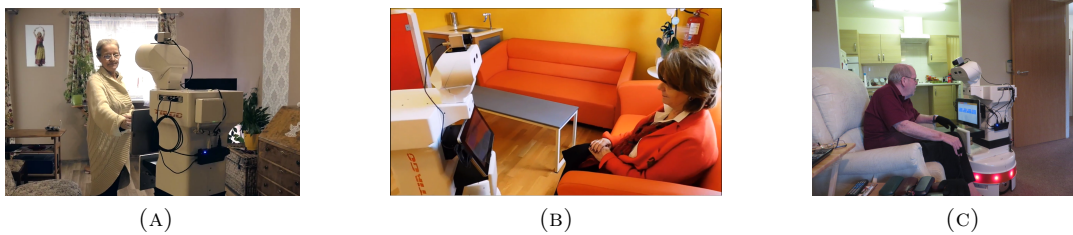


FIGURE 8.16: Participants interacting with the ENRICHME robot; (A) PUMS; (B) AKTIOS; (C) LACE

In order to answer these questions we have recorded how participants used the ENRICHME system GUI in a MongoDB database. The information was recorded using the following keys: user, application name, events (i.e., open or close the application), date, and time. For each cognitive game, we also recorded the game performance, enabling us to track the performance of the users over the time of the experiments.

8.6.1 RQ1: Which is the most used GUI application?

To answer this question, we first have to determine how many times each application was used by each user. Based on the results shown in Table 8.2 the most used GUI application is represented by the Cognitive Games (2723 times), followed by the Environmental data application (491 times). The least used two applications are represented by the Healthy Tips (64 times) and the Call somebody application (94 times).

8.6.2 RQ2: Which is the most played cognitive game?

Table 8.3 shows how many times each game was played by each participant. The most played game is represented by the Hangman game, mostly played by participants P6 (1257 times), P7 (208 times) and P8 (137 times). The second most played game is represented by the Digit Cancellation game, played 272 times by participant P6, and 236 times by participant P8. The two least played games are Integer Matrix game (played 23 times by all 11 participants) and the Speed game (played 24 times by the 11 participants).

8.6.3 Performance analysis

Next, we were interested in finding out the evolution over time of the performance during one cognitive game for one of the participants. For this purpose we selected participant P6 as this participant played more than any of the other participants. For this analysis, we selected the Digit Cancellation game. The rationale behind this decision was that this is the second most played game, and it is not language specific. For Hangman, which is the most played game, there are 150 words from five categories (i.e., Animals, Cities, Vegetables, Fruits, and Objects inside the house). A performance analysis would have to be language specific, taking into consideration the difficulty of the word, the familiarity of the participant with that word, something that is outside the scope of this paper.

In Digit Cancellation the user has 50 seconds to find all occurrences of a randomly chosen digit in a list of random 20 digits (an example is shown in Figure 8.17). The

TABLE 8.2: Usage of each application by the participants

GUI App	P1	P2	P3	P4	P5	P6
Cognitive games	4	4	5	6	93	1843
Health tips	2	4	4	0	13	10
Physical activities	12	22	6	3	6	37
Find object	8	4	12	0	29	4
Environmental data	18	20	22	4	73	125
Agenda	25	33	25	10	24	16
Call somebody	0	26	6	2	9	10
Weather	5	6	4	0	20	41
News	0	7	16	2	7	19

GUI App	P7	P8	P9	P10	P11	Total
Cognitive games	338	418	10	0	2	2723
Health tips	15	14	0	0	2	64
Physical activities	10	35	0	0	2	133
Find object	30	14	0	16	8	125
Environmental data	72	102	16	17	22	491
Agenda	10	21	0	8	5	177
Call somebody	12	20	0	8	1	94
Weather	47	20	0	4	6	153
News	17	19	5	4	7	103

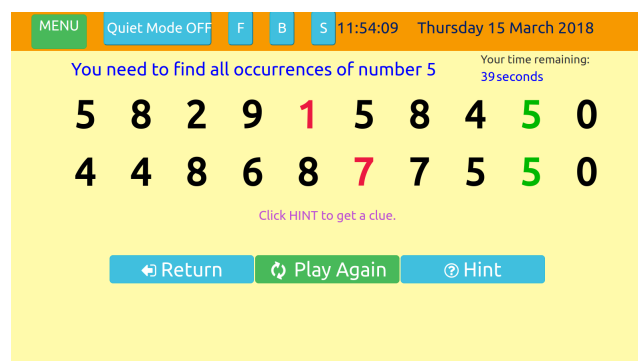


FIGURE 8.17: Example of Digit Cancellation

TABLE 8.3: Number of times each game was played by the participants

Cognitive game	P1	P2	P3	P4	P5	P6
Digit Cancellation	0	0	0	3	18	272
Letter Cancellation	0	1	0	0	10	196
Integer Matrix Task	0	0	0	0	1	13
Decimal Matrix Task	0	4	2	0	0	24
Memory game	0	0	0	0	12	25
Hangman	2	2	2	1	36	1257
Speed game	0	0	0	0	0	24
Puzzle	1	1	1	2	25	12
Stroop Game	1	0	0	0	1	17
Cognitive game	P7	P8	P9	P10	P11	Total
Digit Cancellation	102	236	4	0	1	636
Letter Cancellation	51	23	0	0	0	281
Integer Matrix Task	5	0	4	0	0	23
Decimal Matrix Task	5	0	0	0	0	35
Memory game	0	3	0	0	0	40
Hangman	208	137	0	0	1	1646
Speed Game	0	0	0	0	0	24
Puzzle	14	14	0	0	0	70
Stroop Game	3	5	2	0	0	29

minimum number of occurrences of the chosen digit is 3 and the maximum is 9. The performance parameters of interest are the number of mistakes made and the time needed to find all occurrences. The number of occurrences in each game is random, therefore, we had to select all games that had the same number of occurrences. Participant P6 played the Digit Cancellation game 272 times. There were 132 games with 3 occurrences, 77 games with 4 occurrences, 35 games with 5 occurrences, 17 games with 6 occurrences, 7 games with 7 occurrences, and 3 games with 9 occurrences.

For this analysis we selected the 35 games with 5 occurrences, the 17 games with 6 occurrences and the 7 games with 7 occurrences. We did not consider the games with 3 and 4 occurrences, as the time needed to finish these games was short ($M=5.36$, $M=5.92$, respectively) compared to the 5, 6 and 7 occurrences ($M=8.22$, $M=7$, $M=8.85$, respectively).

As no mistakes were made, only the time needed to find all occurrences is shown in Figure 8.18. The performance is better over time (mostly below 10 seconds), except for games numbers 25 to 28 in case of the games with 5 occurrences, game number 11 for the games with 6 occurrences, and game number 5 for the games with 7 occurrences, when there are peaks in game time. These games were all played in the same day between 13h52 and 13h55. We believe these longer game duration times could be explained by external factors (e.g., fatigue, not feeling well, being distracted by someone else, ...) that cannot be detected by the robot and therefore are not available for further analysis.

We applied Spearman's rank correlation to find out the significance of the improvement in game performance. We obtained the following results: for the games with 5 occurrences ($p=0.05$, $r=-0.32$), for the games with 6 occurrences ($p=0.07$, $r=-0.43$), and for the games with 7 occurrences ($p=0.08$, $r=-0.68$).

Based on these results, we can conclude that a statistically significant decreasing trend ($r=-0.32$) in game times could be observed for games with 5 occurrences. Games with more occurrences, while also displaying a decreasing trend, only approach statistical significance, prompting for further experimentation and analysis before confirming an improvement.

In the next section, we provide a discussion of these results together with some of the lessons that we learned and the limitations of the study.

8.6.4 Discussion

The previously presented results show that all 11 participants interacted with the ENRICHME system. However, there are large differences in the way the participants used the applications provided by the system. Some of them used each application extensively (for example, participants P6 and P8), while others did not use them very much (e.g., participant P9). These differences can be attributed on one part to the preferences of the participants (they might have enjoyed some applications more than others), and on the other part, there was some data that was lost due to technical problems.

The system was tested for a duration of up to 10 weeks. Thus while playing the cognitive games, the participants experience different emotions or moods. As these have an influence on performance, at random times, the self-reported mood of the participants should be recorded.

Another aspect that should be mentioned is that participants might have encountered different difficulties in using the applications. Even if the care givers were trained in how to use each application, sometimes the presence of the technical members is required in order to better explain how a certain application should be used.

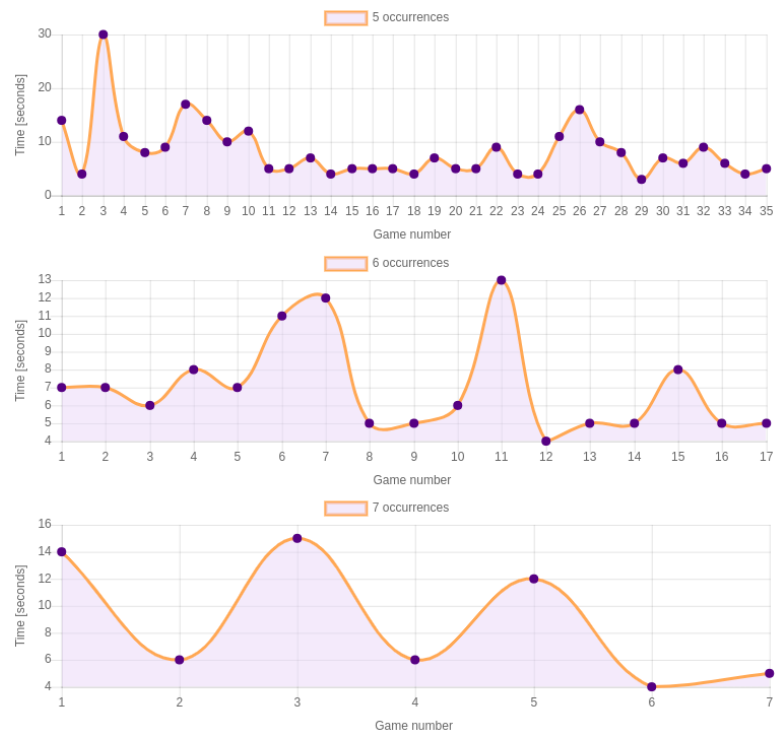


FIGURE 8.18: Performance evolution over time of participant P6 for Digit Cancellation

We believe that it is recommended for the technical partners to be present when the system is presented to the end user. This will ensure that the user knows exactly what to expect from the system.

8.7 Conclusion

In this chapter, we presented some of the results from the validation phase of the H2020 ENRICHME Project as well as the results from a 5-day interaction scenario with one of the ENRICHME end-users. The system was deployed in the houses of 11 participants, who tested the system for a duration of up to 10 weeks. The participants have used all the applications that are part of the system. Some of the most used ones are the cognitive games, and visualizing the data from the environmental sensors (sensors that are placed on the robot and in the house).

8.8 Contribution

The contributions of this chapter consisted in the development of the ENRICHME graphical interface. Based on the advice provided by the health-care professionals and the needs and requirements expressed by the elderly, the nine cognitive games and the physical activities applications were developed for the ENRICHME system. Furthermore, the ENRICHME system was tested for a duration of 5 days with one of the elderly participants.

This work has been published in the International Conference on Social Robotics (ICSR 2016, 2018) (Agrigoroaie et al., 2016; Agrigoroaie et al., 2018b) and the IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2018) (Agrigoroaie et al., 2018c).

Chapter 9

Assistive applications (II)

9.1	Introduction	125
9.2	Literature review	126
9.2.1	Insomnia and cognitive performance	126
9.3	Experimental design	127
9.3.1	Questionnaires	127
9.3.2	Participants	127
9.3.3	Scenario	128
9.3.4	Data recording and analysis	130
9.4	Results	132
9.4.1	Performance results	132
	CPT Task	132
	Integer Matrix Task	134
	Stroop Task	134
	Discussion	134
9.4.2	Performance results for individuals with insomnia	135
	CPT Task	135
	Integer Matrix task	140
	Stroop Task	142
9.4.3	Physiological response analysis	143
	CPT Task	144
	Integer Matrix Task	152
	Stroop Task	154
9.5	Discussion	156
9.6	Conclusions	158
9.7	Contribution	158

9.1 Introduction

Sleep disorders that are responsible for sleep deprivation like insomnia and sleep deprivation have in general numerous consequences on human health and more specifically on the cognitive performances. About 25% of people complain of difficulty falling asleep, and 40% complain of difficulty staying asleep at least a few nights a week (Bonnet et al., 2010; Durmer et al., 2005). The sleep deprivation as a consequences of insomnia or behavioral sleep restriction can lead to numerous cognitive impairments with great impact on general quality of life (e.g., car accidents), and more globally on a significant decrease in cognitive functioning, poor judgment, and decrease in attention (Dinges et al., 1997). It was also already proven that the consequences of sleep deprivation and insomnia could vary with the circadian rhythm and the chronotype and it has been proved that it could be correlated with the internal temperature (Blatter et al., 2007). Despite the fact that hundreds of published studies

exist on the effects of total sleep deprivation, to the best of our knowledge there are few research works focusing on chronic partial sleep restriction as we can find in sleep disorders and especially in insomnia. Moreover, neurocognitive measures vary widely among studies (Durmer et al., 2005). Three categories of measurement commonly used in sleep deprivation studies include cognitive performance, motor performance, and mood, but very few are using sensors like GSR and thermal camera in order to rely neurocognitive task to the physiological aspect.

In the field of Human Computer Interaction (HCI), there are some studies that have investigated the relationship between sleep and different technologies (Rodgers et al., 2016), and how new technologies can encourage healthy sleep patterns (Choe et al., 2011; Aliakseyeu et al., 2011).

In this Chapter, we will focus on the neurocognitive impairments detection. More specifically, the purpose of this preliminary study is to investigate the impact of sleep deprivation and insomnia, morningness-eveningness type (ME-type), and personality, on the cognitive performance during multiple cognitive tasks with circadian component. Therefore, our main research questions are:

- Will individuals without insomnia perform better than individuals with insomnia?
- How does user profile (i.e., gender, age, personality, chronotype) influence performance?
- How will the different physiological parameters vary?

First, we are going to present a literature review for the relationship between insomnia and cognitive performance. Next, the experimental design is detailed. We then present the results of this study, a short discussion and finally conclude this chapter.

9.2 Literature review

For the better understanding of the context in which our work can be placed, we believe there are three main research threads for which a short literature review should be provided. The first one is represented by the relationship between insomnia/sleep deprivation and cognitive performance. The second one is represented by the relationship between ME-type and cognitive performance. And finally the third one is represented by the relationship between individual internal state and cognitive performance. In Chapter 2, it was already presented the relationship between individual internal state and cognitive performance, as well as that between ME-type and cognitive performance.

9.2.1 Insomnia and cognitive performance

It is estimated that around 10% of the general population is suffering from sleep problems severe enough to be diagnosed as insomnia (Ohayon et al., 2009).

The authors of (Fortier-Brochu et al., 2012) have performed a meta-review of the literature dedicated to find the impact of insomnia on daytime cognitive performance. They have reviewed 24 studies that were performed on a total of 639 individuals with insomnia and 558 individuals without insomnia. In the studies, it was found that insomnia has a significant impact on tasks related to problem solving and retention in working memory. No differences were found for general cognitive

function tasks, different dimensions of attention (e.g., speed of information processing, reaction time, sustained attention/vigilance). Their conclusion was that further investigation is needed in order to better understand the impact of insomnia on cognitive performance. But it is now established that short sleeper insomniac patients have more severe consequences as a result of sleep deprivation (Dinges et al., 1997).

The review of the research carried out with the purpose of investigating the relationship between ME-type and cognitive performance has shown that the results differ depending on the cognitive task performed. To the best of our knowledge, there is no study that investigated the relationship between ME-type, time-of-day, and the performance during a continuous performance task. In the field of HCI, the morningness-eveningness was not investigated.

9.3 Experimental design

9.3.1 Questionnaires

Throughout the day of the experiment, each participant was given three questionnaires to fill out. The Morningness-Eveningness Questionnaire (MEQ) and the Eysenck Personality Questionnaire (EPQ) were already presented in Chapter 4.5. Next, the last of the three questionnaires is going to be briefly presented:

Insomnia Severity Index (ISI) (Bastien et al., 2001) represents a screening tool for evaluating insomnia. It is one of the most widely used assessment tools for clinical and observational studies of insomnia. It features seven questions and it can assign individuals to one of the four groups: no clinically significant insomnia, subthreshold insomnia, moderate clinical insomnia, and severe clinical insomnia. For this study, we divided participants into two groups: clinical insomnia (moderate and severe insomnia) and non-clinical insomnia (subthreshold insomnia and no clinically significant insomnia).

For the purposes of this study, based on the results of the MEQ, we consider participants either morning or evening type.

9.3.2 Participants

For this study, a total of 15 patients (9 males and 6 females) from the Sleep Disorder Unit of the University Hospitals Pitié Salpêtrière - Charles Foix, located in Paris, agreed to perform the cognitive tasks presented in Section 9.3.3 and signed a consent form. Their ages ranged between 11 and 65, with a mean of $M = 45.2$, and a standard deviation of $SD = 14.97$. Based on the age, we divided the participants into three groups. Participants with ages up to 35 were considered in the **young/early adulthood** age group. Participants with ages between 35 and 55 were considered in the **middle adulthood** age group, while participants with ages greater than 55 were considered in the **late adulthood/old** age group. Our participant sample is not homogeneous according to age. This is a preliminary study that evaluates the differences in the variation of the physiological parameters between the four cognitive tasks as well as to investigate the performance during the four cognitive tasks.

Two of the participants are retired, one is a child still in school, while the others are all active in the workforce. Their occupations cover a broad spectrum, e.g., bank technician, manager, architect, heavyweight truck driver, and lawyer. There are three participants that smoke on a regular basis. Seven of the 15 participants drink coffee everyday, while the other 8 either drink tea or do not drink any caffeinated

TABLE 9.1: Distribution of participants

Gender		Age group		ME-type	
Males	9	Early adulthood	4	Evening	6
Females	6	Middle adulthood	6	Morning	8
		Late adulthood	5		
Extroversion		Neuroticism		Insomnia	
Intro	6	Low	6	Clinical	10
Extro	8	High	8	Non-clinical	4
Sleep time		Sleep quality		Medication	
Short	8	Medium	5	Yes	8
Normal	7	Bad	7	No	7

drinks. Table 9.1 shows a distribution of the participants based on their background information, the results of the questionnaires, and some of the sleep measures.

9.3.3 Scenario

The experiment presented in this study took place in the Sleep Disorder Unit of the University Hospitals Pitié Salpêtrière - Charles Foix. The participants recruited for this study are all outpatients of the Sleep Disorder Unit that were admitted with either sleep or attention problems. For the patients suffering of sleep disorders, the experimental procedure started one week before the experiment day. First, they were given a watch containing an actimeter in order to detect sleep disorders and circadian disorder (e.g., delayed sleep phase syndrome). The watch was equipped with a 3D accelerometer with a specific software that is used to quantify the intensity and duration of daily physical activity. Then, based on the recorded data, it can identify irregular activity patterns of the assessment of sleep quality with the final goal of measuring the sleep patterns of the patients. During the night before the experiment, almost all participants underwent a home overnight polysomnography (PSG).

The PSG used for this study is a combination of electroencephalography (EEG), in which 6 electrodes are attached on the patient's scalp in order to record his/her brain wave activity, and 2 continuous electro-oculography (EOG), which record eye movement. The air flow through the patient's nose and mouth are measured by heat-sensitive devices (i.e., thermistors). This can help detect episodes of apnea (stopped breathing), or hypopnea (inadequate breathing). Another test called pulse oximetry measures the amount of oxygen in the blood. The electrical activity of the patient's heart is also measured on an electrocardiogram (ECG).

This enabled the medical practitioners to determine the sleep time and the sleep quality of the night before the experiment. On the day of the experiment, each patient arrived at the hospital early in the morning (i.e., at around 8h30). A series of four experimental sessions were planned throughout the day, at intervals of two hours. Table 9.2 shows an overview of the experimental tasks performed during each session. The experimental tasks that the participants had to perform are:

- Maintenance of Wakefulness Test (MWT)
- Continuous Performance Test (CPT)

TABLE 9.2: Tasks performed by participants

Session	Approximate start time	Tasks
Session #1	8h45 - 9h00	MWT CPT Cognitive Games
Session #2	10h45 - 11h00	MWT
Session #3	12h45 - 13h00	MWT CPT Cognitive Games
Session #4	14h45 - 15h00	MWT

- Cognitive games: Integer Matrix Task, and Strop Test

The MWT and the CPT represent two of the standard tests used by the Sleep Unit in order to better understand the influence of sleep deprivation on the ability of the sleep deprived individuals to stay awake for a given amount of time and to measure sustained attention. The third task, the cognitive games, are the ones developed by us that we wanted to test with a different population. Next, we provide a short description for each test.

Maintenance of Wakefulness Test (MWT) (Mitler, 2004): is a test used to determine if an individual is able to stay awake for a given amount of time. The MWT is used here to figure out if despite the sleep deprivation, the patient keep an hyperarousal typical of insomniac patients. In our experiment, it allows to monitor the wakefulness and avoid a nap before the cognitive task. The task took place in a hospital room. The participants had to stay awake for a duration of 40 minutes with only a dim light on and without any other distractions. During this time a technician recorded the EEG. If during the test the patients fell asleep, the test was terminated and the next task was performed.

Continuous Performance Test (CPT): (Conners et al., 2000) was developed to measure sustained attention and learning disorders. The test is performed on a computer and it lasts for 14 minutes. A series of letters are displayed on the screen, each letter being shown for a duration of 250 milliseconds. The respondent has to press on the space bar every time a letter appears *except* when the letter "X" is displayed. The letter displayed is white and it is shown on a black background. There are three inter-stimulus intervals (ISIs): 1 second, 2 seconds and 4 seconds. The test is made up of six main blocks, with three subblocks (one for each ISI) each containing 20 trials (letter presentations). During each subblock the letter "X" is displayed two times. The letters and their orders are chosen randomly. The order of the different ISIs is randomized between blocks.

The CPT software provides the score for the test and a report with the main measures recorded. The report contains a summary of the overall measures (the values, t-scores, and percentile), the data for each block, and the data for each subblock. For the overall measures, the t-scores and the percentiles represent the score of the individual taking the current test relative to the population (i.e., same gender and age group) average.

Some of the measures provided are: the hitRT (i.e., the time needed to press on the



FIGURE 9.1: Experimental setup

space bar for a letter that is not "X"), number of omissions (i.e., when a response is not given when a letter that is not "X" is shown), the number of commissions (i.e., when a response is given when the "X" letter is shown).

Cognitive games: the Stroop game and the Integer Matrix task (medium difficulty level - matrix size of 3x3, difficult difficulty level - matrix size of 4x4), presented in Chapter 3 were used for this study. The two cognitive games were performed on a touch-screen, placed on the same table as the computer for the CPT test (as shown in Figure 9.1). Both games were translated into French.

For the purposes of this study, of interest are Session #1 and Session #3, as during these two sessions the participants performed cognitive tasks (as shown in Table 9.2). In the following section, we present the data that was recorded during these two sessions.

9.3.4 Data recording and analysis

For Session #1 and Session #3, we recorded both task performance for the two cognitive tasks (i.e., CPT and cognitive games) and the physiological data. The physiological data was recorded with three sensors: a thermal camera (i.e., Optris PI640), an RGB-D camera (i.e., Asus Xtion Live Pro), and a GSR sensor (i.e., Grove GSR).

Task performance for CPT

For the CPT there are three types of results: overall results, block results and sub-block results. For this study, we considered for the analysis the subblock measures. We believe that they enable us to better determine the performance than using the overall results. Therefore, the parameters used to assess task performance are: the hitRT, the number of omissions and the number of commissions. According to (Conners et al., 2000), all reaction times that are less than 100 ms should be excluded, as due to physical limitations, we are not able to be that fast, unless we press on the space bar without waiting for the stimulus (i.e., anticipation response). Therefore, we excluded from our analysis all reaction times that are less than 100ms.

Task performance for cognitive games

During the Integer Matrix task, the participants were careful not to make any mistakes. Therefore, the number of mistakes was very low. Out of a total of 894 matrices solved in the medium level, there were only 4 mistakes (0.45%). For the difficult level, out of a total of 594 matrices solved there were 12 mistakes (2.02%). As a result, the

number of mistakes made by the participants was not considered for analysis. For the statistical analysis, we used the number of correct answers and the average time to solve a matrix, for each of the two difficulty levels.

For the Stroop Task, there were a total of 1200 trials (40 trials during each game). There were 12 mistakes (1%) and 90 trials for which there was no answer (7.5%). The percentage of mistakes is quite low compared to the number of correct answers. Therefore, we considered for the analysis the number of correct answers, the number of trials without any answer, and the average reaction time.

Physiological measures

The sensors used for recording the physiological parameters (i.e., RGB, thermal and GSR) were placed as shown in Figure 9.1. The RGB and thermal cameras were placed in front of the participant, while the GSR sensor was placed on the non-dominant hand of each participant (i.e., on the left hand).

From the thermal data we extracted and analyzed the variation of the facial temperature in four regions of interest (ROI): the forehead, the periorbital region, the tip of the nose, and the perinasal region. From each ROI, the following parameters were extracted and used for analysis: the average temperature, the temperature range (i.e., maximum - minimum temperature), and the rate of change of the temperature (i.e., slope).

From the GSR data, we extracted the accumulative GSR (AccGSR), the number of peaks and the average GSR value. The authors of (Nourbakhsh et al., 2013; Setz et al., 2010) have shown that AccGSR is a good indicator of the cognitive load. We performed the calibration for each feature as suggested by (Nourbakhsh et al., 2013) as the GSR values are highly subjective. More specifically, we took the feature for each task for each participant and we divided it by the average of all the similar features of all the tasks performed by that participant.

From the RGB data we extracted the number of blinks by using the method presented in Chapter 4.3.

The data from the three sensors were recorded during all three tasks (i.e., the MWT, the CPT, and the cognitive games). For this study we only analyse the variation of the physiological parameters during the two cognitive tasks: the CPT and the cognitive games.

Other measures

As previously presented, for each participant we also measured their sleep time and their sleep quality from the night before the day of the experiment. The sleep time was recorded in minutes. We divided the participants into two groups depending on the sleep time. The first group is the short sleep time group with a total sleep time of less than 360 minutes (i.e., 6 hours). The maximum sleep time for our participants was of 483 minutes (i.e., 8 hours), therefore, all sleep times greater than 360 minutes were categorized as normal sleep time.

For the sleep quality, the analysis of the polysomnography was conducted. The quality was determined by the cross-analysis of 3 parameters: the micro-arousal index, the time awake during the sleep, and the Slow-wave Sleep (SWS) and Rapid Eye Movement (REM) percentage. If the 3 parameters show more than 20 microarousals per hour, more than 15% of wake time during the sleep and less than 80 minutes of SWS and 20% of REM sleep - the sleep quality was quoted as BAD; if one of these factor was normal - the sleep quality was quoted as MEDIUM; and if all parameters were normal - the sleep quality was quoted as GOOD.

9.4 Results

In this section, we are going to present the results of this study. This section is divided into three main parts, each corresponding to a research direction. First, we are going to present the task performance results for all 15 participants. In the second part, we analyse the performance of the individuals diagnosed with insomnia. And in the third part, we perform the analysis for the physiological response during the cognitive tasks. For the last part, we first investigate the variation of each physiological parameter, then, we investigate the correlations between the performance measures and the physiological parameters, as well as that between the different physiological parameters.

For the first two parts we consider as factors: the gender, the age group (i.e., young adulthood, middle adulthood, and late adulthood), if they took any sleep medication, their neuroticism and extroversion level, the ME-type (i.e., morning or evening), the session (i.e., morning or afternoon), and the information about their sleep (i.e., sleep time and sleep quality). For each part we are going to present the results for each of the four tasks performed by the participants: CPT task, Stroop Task, Integer Matrix Task - Medium level, and Integer Matrix Task - Difficult level.

For each analysis, we first test if the investigated parameter is normally distributed by applying a Shapiro-Wilk normality test. For the normally distributed parameters we apply an ANOVA test, while for the non-normally distributed data we apply a Kruskal-Wallis test. To determine differences between more than two groups we apply pairwise t-tests (for normally distributed data) or the Wilcoxon test (for the non-normally distributed data), both with a Bonferroni correction.

9.4.1 Performance results

CPT Task

The performance measures for the CPT task are the number of omissions (i.e., when a response is not given when a letter that is not “X” is shown), number of commissions (i.e., when a response is given when the letter “X” is shown), and the hitRT (i.e., the time needed to press on the space bar for a letter that is not “X” when a correct answer is given).

We investigated both the block and the subblock results.

Block results

The number of omissions are not normally distributed, therefore, we apply a Kruskal-Wallis rank sum test. The significant factors found are: the gender ($\chi^2 = 8.15$, $p = 0.004$), the age group ($\chi^2 = 15.37$, $p = 0.0004$), the sleep time ($\chi^2 = 18.08$, $p < 0.0001$), and if the participants took any sleep medication ($\chi^2 = 12.07$, $p = 0.0005$).

We found that males have a higher number of omissions than females ($M_{males} = 2.01$, $M_{females} = 0.31$). The middle adulthood group of participants had a lower number of omissions ($M_{middle\ adulthood} = 0.27$) compared to the young adulthood group ($p = 0.0037$, $M_{young\ adulthood} = 3.52$), and compared to the late adulthood group as well ($p = 0.0006$, $M_{late\ adulthood} = 0.86$). Individuals with short sleep time had a significantly higher number of omissions than the individuals with normal sleep time ($M_{normal} = 0.26$, $M_{short} = 2.28$).

Next, we found that the number of commissions is also influenced by the age group of the participants (see also Figure 9.2). As with the number of omissions, the young adulthood group had the most commissions ($M_{young\ adulthood} = 2.66$), and significantly more commissions than the middle adulthood age group ($p < 0.0001$, $M_{middle\ adulthood} = 1.48$), and more commissions than the late adulthood age group

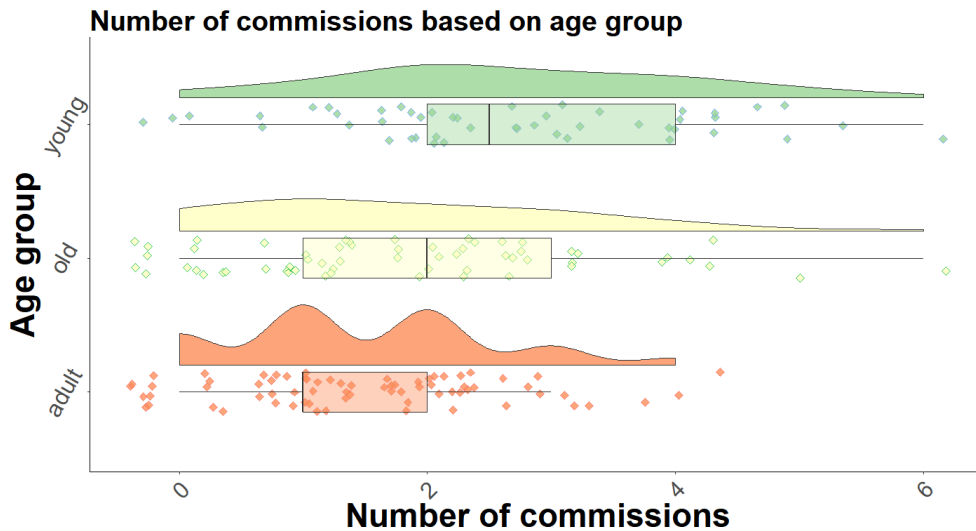


FIGURE 9.2: Number of commissions based on the age group of the participants

($p = 0.003$, $M_{late\ adulthood} = 1.78$). Moreover, the participants in the young adulthood age group were significantly faster than the participants in the middle adulthood age group ($p = 0.001$, $M_{young\ adulthood} = 390$, $M_{middle\ adulthood} = 418$), and they were also significantly faster than the late adulthood age group ($p = 0.0001$, $M_{late\ adulthood} = 424$). Therefore, individuals in the young adulthood age group are faster than the other two age groups, but they also make more commissions and have more omissions as well.

Based on the ME-type of the participants, evening type individuals made significantly more commissions than morning type individuals ($\chi^2 = 32.15$, $p < 0.0001$, $M_{morning} = 1.32$, $M_{evening} = 2.52$). Moreover, evening type individuals are also significantly faster than the morning type individuals ($F(1, 166) = 41.02$, $p < 0.0001$). The average speed difference between the two groups is of 40ms. Therefore, evening type individuals are faster, but they also make more mistakes.

Subblock results

When looking at the subblock results, we found that males had a higher number of omissions than females ($\chi^2 = 12.04$, $p = 0.0005$, $M_{males} = 0.67$, $M_{females} = 0.09$). Based on the age group of the participants, middle adulthood participants had the least number of omissions, which is significantly less than the young adulthood group ($p = 0.0004$, $M_{middle\ adulthood} = 0.09$, $M_{young\ adulthood} = 1.15$), and also significantly less than the late adulthood age group ($p = 0.0001$, $M_{late\ adulthood} = 0.28$). Individuals with short sleep time had a significantly higher number of omissions than the individuals with normal sleep time ($\chi^2 = 22.43$, $p < 0.0001$, $M_{short} = 0.76$, $M_{normal} = 0.07$). As can be seen, the same results were found for the block measures as well.

For the number of commissions, the significant factors found are: the age group ($\chi^2 = 22.57$, $p < 0.0001$), the ME-type ($\chi^2 = 40.77$, $p < 0.0001$) and the extroversion level ($\chi^2 = 24.7$, $p < 0.0001$).

Introverted participants made almost twice as many commissions than extroverted participants ($M_{intro} = 0.78$, $M_{extro} = 0.47$), having a faster reaction time than extroverted participants ($\chi^2 = 33.56$, $M_{intro} = 392$, $M_{extro} = 416$).

Evening type individuals made more commissions than morning type individuals

($M_{morning} = 0.44$, $M_{evening} = 0.83$), and they were also faster than morning type individuals ($\chi^2 = 63.7$, $p < 0.0001$, $M_{evening} = 380$, $M_{morning} = 426$).

Young adulthood participants made the most commissions compared to the late adulthood participants ($p = 0.002$, $M_{young\ adulthood} = 0.86$, $M_{late\ adulthood} = 0.59$), and the middle adulthood participants ($p < 0.0001$, $M_{middle\ adulthood} = 0.49$). Furthermore, the young adulthood age group participants were also the fastest when compared to the late adulthood age group participants ($p < 0.0001$, $M_{young\ adulthood} = 379$, $M_{late\ adulthood} = 423$) and when compared to the middle adulthood age group participants ($p < 0.0001$, $M_{middle\ adulthood} = 417$).

The results for both Block measures and Subblock measures are very similar. The participants in the young adulthood age group are fast, but they also make more mistakes than the other participants, which is a sign of impulsivity. Introverted participants are also fast but make more mistakes, as well as evening type individuals.

Integer Matrix Task

As previously shown, the performance measures for the Integer Matrix Task (for both difficulty levels) are the number of correctly solved matrices, and the average time needed to solve a matrix.

For the medium level of this task, we did not find any significant results for the task performance based on the different demographic and personality factors.

On the other hand, for the difficult level of this task, we found that the average time needed to solve a matrix is influenced by the sleep time ($\chi^2 = 4.32$, $p = 0.03$), and if the participants took any sleep medication ($\chi^2 = 4.67$, $p = 0.03$).

Concerning the sleep time, individuals with short sleep time needed more time to solve a matrix than the individuals with a normal sleep time ($M_{short} = 7.94$, $M_{normal} = 6.89$). Moreover, the individuals who took sleep medication took on average longer to solve a matrix than the individuals who did not take the sleep medication ($M_{withmedication} = 7.96$, $M_{withoutmedication} = 6.87$). No significant results were found for the number of correctly solved matrices.

Stroop Task

The performance measures for the Stroop Task are the average time to give an answer, the number of correct answers, and the number of trials for which no answer was given.

We first considered the average time to give an answer (RTTotal). The first significant factor that we found is the age group of the participants ($F(2, 27) = 12.86$, $p = 0.0001$). This results is also graphically represented in Figure 9.3. A pairwise comparisons test (i.e., t-test) with Bonferroni correction showed that late adulthood individuals are slower than young adulthood individuals ($p = 0.0001$, $M_{young\ adulthood} = 1.64$, $M_{late\ adulthood} = 2.12$), and they are also slower than the middle adulthood age group individuals ($p = 0.0049$, $M_{middle\ adulthood} = 1.83$).

We also found that morning type individuals are slower than evening type individuals ($F(1, 26) = 11.81$, $p = 0.001$, $M_{morning} = 2.01$, $M_{evening} = 1.70$).

These results are in accordance with the results found for the CPT task.

Discussion

In this section we have performed the performance analysis for the all 15 participants who took part in this study. For both the CPT task and the Stroop Task, we found that evening type individuals are faster than morning type individuals. This translates to a higher number of mistakes made during the CPT task. Moreover,

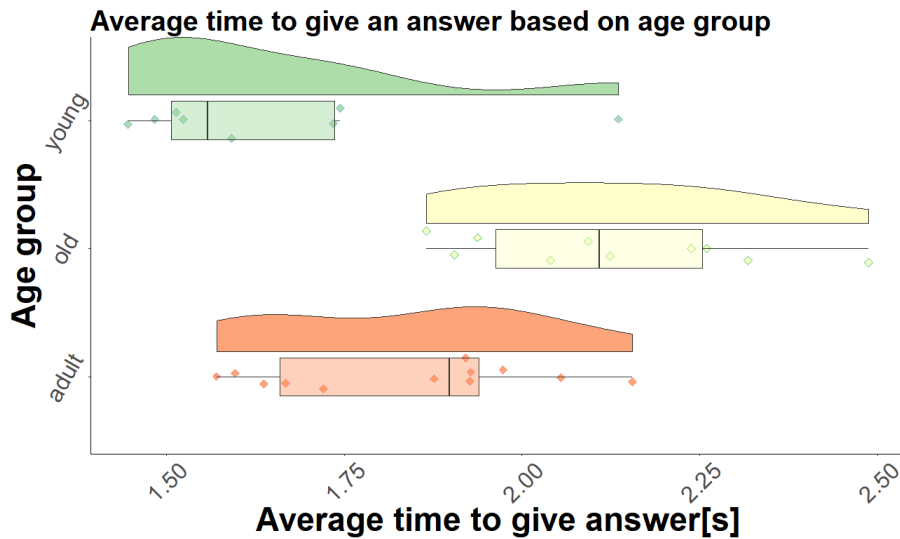


FIGURE 9.3: Average time to give an answer (RTTotal) based on the age group for the Stroop Task

our results show that young adulthood individuals have a tendency to be very fast (for both CPT task and the Stroop task), but they also make more mistakes (for the CPT task). Sleep time influences the performance both during the CPT task and during the difficult level of the Integer Matrix task, with more omissions being made by the individuals with short sleep time, and a longer time needed to solve a matrix than for the individuals with a normal sleep time. These results suggest that both young adulthood individuals, as well as evening type individuals have a behavior associated with impulsiveness. No psychological questionnaire was used to measure the impulsiveness of the participants, therefore, we cannot confirm that indeed these participants have high levels for the impulsiveness scale.

Next, we are going to investigate the results for the individuals with clinical insomnia.

9.4.2 Performance results for individuals with insomnia

CPT Task

The parameters used to characterize the cognitive performance are: the average time needed to answer correctly (hitRT), the number of omissions, and the number of commissions. Before performing the analysis, a normality test was applied on the dependent variables. None of the variables are normally distributed. Therefore, we tested each dependent variable by using non-parametric tests.

Next, we present the results for each research question. Our first research question is related to the performance of individuals depending on their insomnia level. Out of the total participants, 10 have clinical insomnia, 4 do not have clinical insomnia, and for one participant we do not have this result (i.e., our youngest participant). As the two groups are not of equal sizes, we applied three Mann-Whitney tests, one for each of the dependent variables (i.e., hitRT, number of omissions, and number of commissions) characterizing the cognitive performance. Our results show that individuals with insomnia are significantly faster than individuals without insomnia ($W = 18980, p < 0.0001$) (see Figure 9.4).

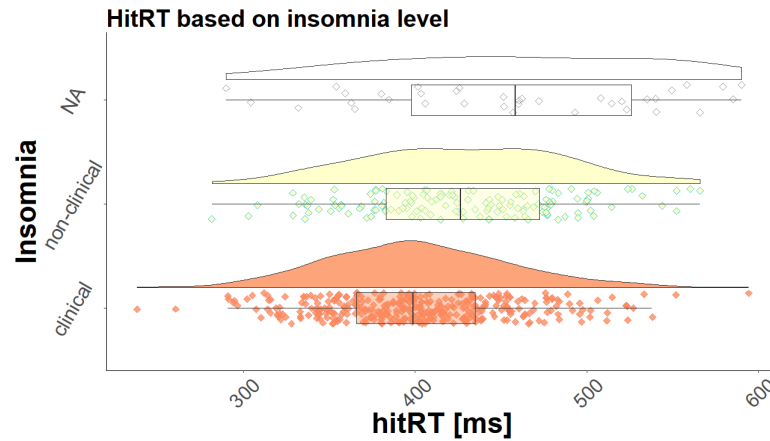


FIGURE 9.4: HitRT based on presence of insomnia

TABLE 9.3: Distribution of individuals with insomnia

Gender		Age group		ME-type	
Males	5	Early adulthood	3	Evening	6
Females	5	Middle adulthood	3	Morning	4
		Late adulthood	4		
Extroversion		Neuroticism		Sleep time	
Intro	4	Low	3	Short	6
Extro	6	High	7	Normal	4
Sleep quality		Medium		4	
		Bad		6	

When considering the number of commissions and the number of omissions, we found that individuals with insomnia commit more omissions ($W = 27870, p = 0.02$) and more commissions ($W = 30374, p = 0.0008$) than individuals without insomnia.

We can conclude that individuals with insomnia are faster than individuals without insomnia. However, the task performance is not measured only by looking at speed. The participants also need to solve the task without making mistakes. Which is not what happened for our participants. Individuals with insomnia committed more omissions and more commissions than individuals without insomnia. Therefore, we can say that individuals without insomnia had a better overall performance than individuals with insomnia. However, due to the small group of participants (4 participants without insomnia), our result cannot be generalized since it may not capture the variation in the broader population.

As the number of individuals without insomnia is quite low (i.e., 4 participants), for our second research question we are only investigating the influence of the user profile for the individuals suffering of insomnia. The factors that determine the user profile are: age, gender, ME-type, extroversion, and neuroticism. The distribution of the participants based on these factors is shown in Table 9.3.

First, we tested the normality of the dependent variables. Only the hitRT is normally distributed ($W = 0.99, p = 0.25$), and therefore, we applied the analysis of variance (ANOVA) for the hitRT, and the Kruskal-Wallis Rank Sum Test for the number of omissions and the number of commissions.

When considering gender as factor, we found that males committed significantly more omissions than females ($\chi^2 = 21.67, p < 0.0001$). On average, male participants had a mean of $M = 1.11$ omissions ($SD = 3.53$), while female participants had a mean of $M = 0.07$ omissions ($SD = 0.29$).

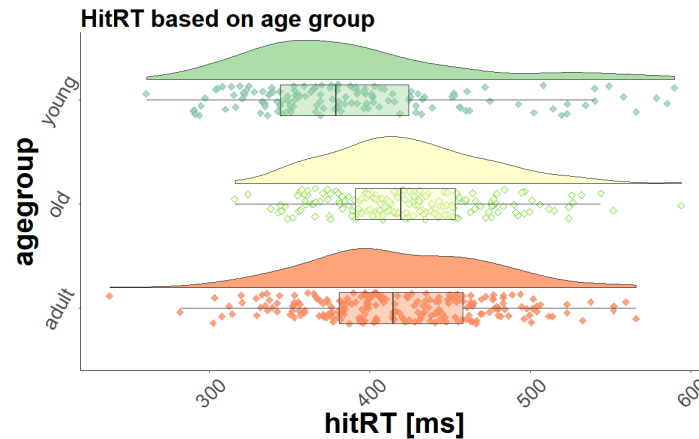


FIGURE 9.5: HitRT based on age group

Age is a significant factor for all three parameters of performance. For hitRT the ANOVA analysis showed a significant result with $F(2, 353) = 36.45, p < 0.0001$, see also Figure 9.5. A pairwise comparisons using t-tests with a Bonferroni correction showed significant differences between young/early adulthood individuals and middle adulthood individuals ($p < 0.0001$), and between young/early adulthood individuals and late adulthood individuals ($p < 0.0001$). Young/early adulthood individuals are significantly faster than the other two groups. No differences were found in speed between the middle adulthood and late adulthood individuals.

For omissions, we also found that age is a significant factor ($\chi^2 = 12.34, p = 0.002$). A pairwise comparisons using Wilcoxon rank sum test with a Bonferroni correction revealed significant differences between the middle and late adulthood groups ($p = 0.0029$), and between the middle and young/early adulthood groups ($p = 0.0031$). The average number of omissions for each group is as follows: young/early adulthood ($M=1.45, SD=4.44$), middle adulthood ($M=0.11, SD=0.63$), late adulthood ($M=0.31, SD=0.73$). As seen from these results, young/early adulthood individuals are significantly faster than the other groups. However, they commit more omissions than the other two groups.

As the number of commissions is also a significant factor ($\chi^2 = 11.14, p = 0.003$), we applied a pairwise comparison using a Wilcoxon rank sum test and a Bonferroni correction. We found significant differences between the young/early adulthood and late adulthood groups ($p = 0.01$), and between the middle and late adulthood groups ($p = 0.012$). The results show that late adulthood individuals commit significantly less commissions than the other two groups.

According to (Conners et al., 2000), age has an effect on the number of omissions, the number of commissions, and on the hitRT. It was found that for the normative population used for developing the tests, young individuals are faster than the adults or the individuals with ages 55+. Our results are in accordance with the results used for developing the test.

For the ME-type factor, we found that evening type individuals are significantly faster than morning type individuals ($F(1, 354) = 46.46, p < 0.0001$), shown also in Figure 9.6. However, even though they are faster, evening type individuals commit significantly more commissions (i.e., response when “X” appears) than morning type individuals ($\chi^2 = 28.56, p < 0.0001$).

When considering the two personality traits (i.e., extroversion and neuroticism), we found that extroverted individuals are faster than introverted individuals ($F(1, 354) =$

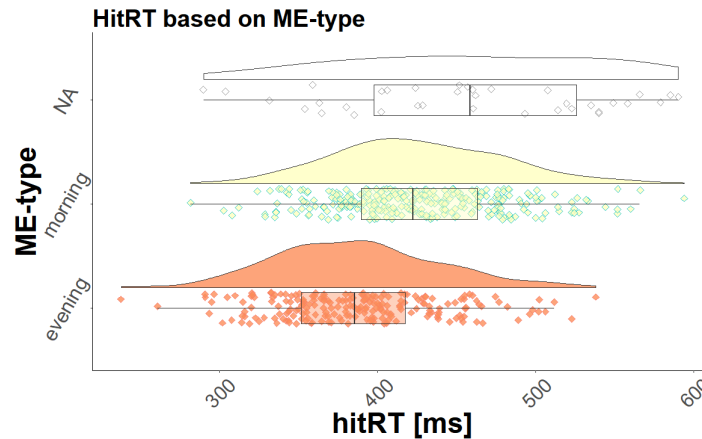


FIGURE 9.6: HitRT based on ME-Type

18.2, $p < 0.0001$) but they also commit more commissions than introverted individuals ($\chi^2 = 13.531, p = 0.0002$).

Individuals with high scores for neuroticism are significantly faster than individuals with low neuroticism scores ($F(1, 354) = 6.04, p = 0.014$). However, they commit more omissions ($\chi^2 = 5.55, p = 0.018$). In contrast, they commit less commissions than individuals with low neuroticism levels ($\chi^2 = 7.06, p = 0.0078$). This result shows that individuals with high neuroticism levels have a tendency to be more impulsive.

Next, we investigated the influence of other factors: sleep time, sleep quality, if the participants took any sleep medication, and the interaction time (i.e., morning or afternoon).

The sleep quality had a significant influence only on the hitRT, with individuals with a medium sleep quality being significantly faster than individuals with a bad sleep quality ($F(1, 354) = 7.97, p = 0.005$). This result is in accordance with the literature (e.g., (Durmer et al., 2005)), which shows that sleep deprivation leads to the diminution of response time.

For the sleep time, we found significant differences for the hitRT ($F(1, 354) = 5.96, p = 0.015$) and the number of omissions ($\chi^2 = 25.37, p < 0.0001$). Individuals with a short sleep time were significantly faster than the individuals with a normal sleep time and they also committed more omissions. Again, this result is in accordance with the literature related to the number of commissions and it can be interpreted as impulsivity or as diminished behavioral inhibition abilities (Cheson et al., 2007).

The sleep medication influenced the number of omissions ($\chi^2 = 12.85, p = 0.0003$) and the number of commissions ($\chi^2 = 7.85, p = 0.005$). The individuals that took the sleep medication committed more omissions and more commissions. These results again are coherent with the literature on hypnotics (i.e., sleeping pills). Indeed, hypnotics impaired performance on serial reaction time (SRT) with increase omissions (Paul et al., 2003).

Lastly, for the CPT, we investigated two interactions: (1) the interaction between the sleep quality and the sleep time and its influence on the task performance and (2) the interaction between the ME-type and the interaction time (i.e., morning or afternoon) on the task performance.

Out of the 10 participants with insomnia, four had a medium sleep quality the night before the experiment, and six had a bad sleep quality. For the medium sleep quality group, half of them had a normal sleep time, and the other half had a short

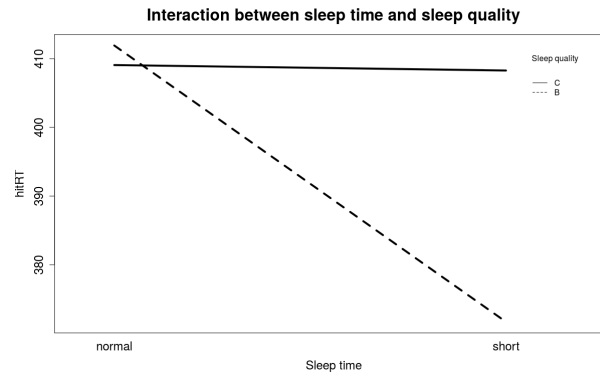


FIGURE 9.7: Interaction between sleep time and sleep quality on hitRT (C=Bad sleep quality; B=Medium sleep quality)

sleep time. For the group with bad sleep quality, two of them had normal sleep time and four had short sleep time.

Figure 9.7 shows the interaction plot between the sleep time and sleep quality and its effect on hitRT. As it can be seen, for a bad sleep quality, the sleep time does not influence speed. The average hitRT for the individuals with bad sleep quality and normal sleep time is $M_{normal} = 409.07$, while for the individuals with a short sleep time is $M_{short} = 408.28$.

On the other hand, for the individuals with a medium sleep quality, the hitRT changes depending on the sleep time. For the group with individuals with normal sleep time the average hitRT is $M_{normal} = 411.92$, while for the group with short sleep time the average hitRT is $M_{short} = 371.69$. A Welch Two Sample t-test revealed a significance level of $p < 0.0001$ between these two groups.

Therefore, we can conclude that, for a bad sleep quality there is no difference in speed depending on the sleep time. While for a medium sleep quality, with a normal sleep time the speed is significantly lower than for a short sleep time. To better understand how this result influences overall performance, we also looked at the effect of the interaction between the sleep time and sleep quality on the number of omissions and the number of commissions for the individuals who had a bad sleep quality. A Welch Two Sample t-test revealed a significance level of $p = 0.0009$ for the number of omissions, with individuals with a short sleep time committing more omissions $M_{short} = 2.22$ than individuals with a normal sleep time $M_{normal} = 0.07$. A significant result was found for the number of commissions as well $p = 0.002$, with individuals with a normal sleep time committing less commissions $M_{normal} = 0.51$ than individuals with a short sleep time $M_{short} = 0.89$.

Therefore, we can conclude that, sleep time does not influence in any way performance if the quality of the sleep was bad. However, for a medium sleep quality, a short sleep time leads to faster reaction time but also to more omissions and more commissions than a normal sleep time.

For the second interaction, we considered the ME-type and the interaction time (i.e., morning or afternoon). Figure 9.8 depicts the interaction plot for the effect of the interaction between the ME-type and the time when the task took place.

As it can be seen, for the morning type individuals there are no differences in the number of omissions between the two sessions ($p = 0.24$). For the evening type individuals, the number of omissions depends on when the task was performed ($p = 0.0019$). During the first session (i.e., morning), the evening type individuals committed less omissions ($M_{morning} = 0.13$), than during the afternoon session

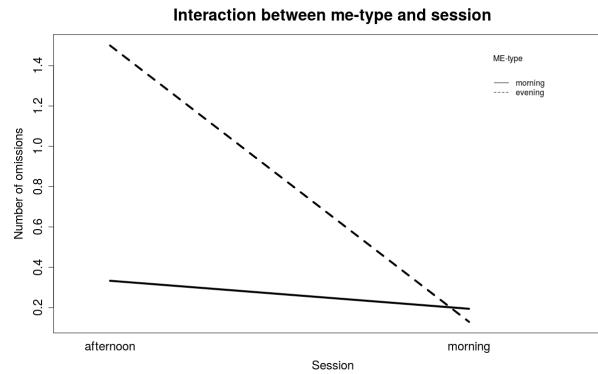


FIGURE 9.8: Interaction between ME-type and interaction time on omissions

($M_{afternoon} = 1.5$). No significant differences were found for neither the number of commissions, nor the hitRT.

Therefore, we can conclude that the user profile does influence task performance for the individuals with insomnia. All factors considered (i.e., age, gender, ME-type, extroversion and neuroticism) have impact on the task performance. Evening type individuals are faster than morning type individuals; males commit more omissions than females; individuals with high neuroticism level are faster and make more omissions than individuals with low neuroticism level; and young/early adulthood individuals are faster but more careless than the individuals that are in the middle adulthood or late adulthood groups.

Integer Matrix task

For the Integer Matrix task, we had two difficulty levels (i.e., medium, and difficult). For both difficulty levels we consider as performance parameters the total number of matrices correctly solved and the average time needed to solve a matrix (RTTotal).

As with the CPT, we first investigated if there are any differences in performance depending if the individual has or not clinical insomnia. We found no significant differences in performance depending on insomnia for neither of the two difficulty levels. We found that for both difficulty levels, individuals without insomnia were faster and as a result they solved more matrices than individuals with insomnia. However, these results are not significant. As the number of individuals without insomnia is lower than the individuals with insomnia, for our second research question we considered for analysis only the performance of the individuals with insomnia. Our main results are summarized below.

For gender, we found significant results only for the difficult level. Female participants were faster than the male participants ($\chi^2 = 4.16, p = 0.04$), and as a result they solved more matrices ($F(1, 18) = 3.85, p = 0.06$).

The other significant factor is the personality trait of neuroticism. Individuals with high neuroticism level were significantly faster than individuals with low neuroticism level for both medium ($\chi^2 = 9.82, p = 0.001$), and the difficult level ($\chi^2 = 9.31, p = 0.002$). As a consequence, they also solved correctly significantly more matrices in both the medium ($F(1, 18) = 32.48, p < 0.0001$) and the difficult level ($F(1, 18) = 35.46, p < 0.0001$). Figure 9.9 shows the number of correctly solved matrices depending on the neuroticism level and the difficulty level, while Figure

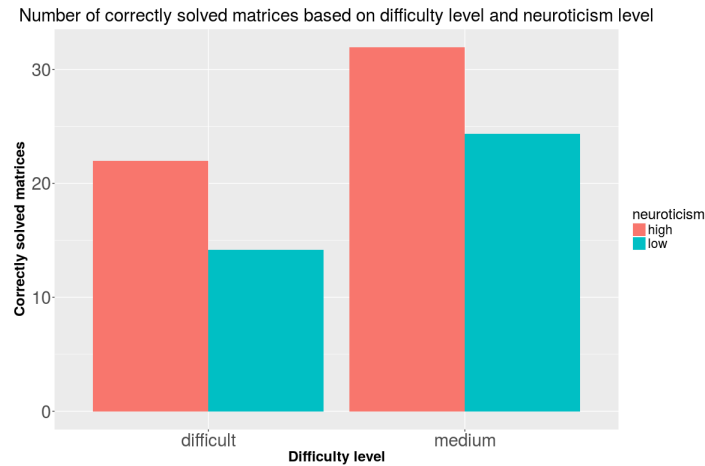


FIGURE 9.9: Correctly solved matrices based on difficulty and neuroticism level

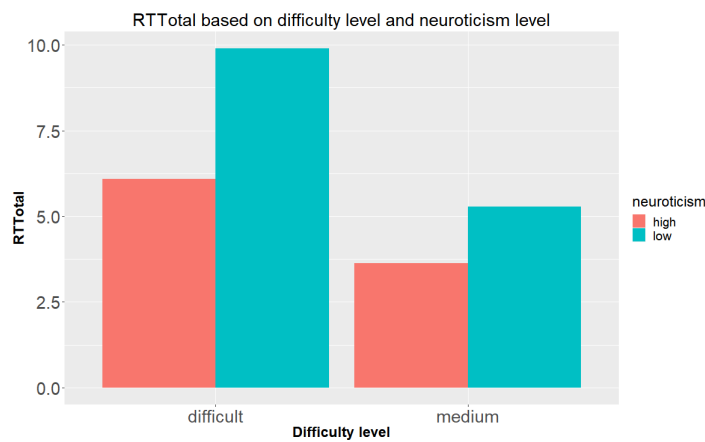


FIGURE 9.10: Time to solve a matrix based on difficulty and neuroticism level

9.10 shows the time needed to solve a matrix based on the neuroticism level and the difficulty.

Concerning the sleep time factor, we found that it influences only the performance during the difficult level. Individuals with normal sleep time were significantly faster than the individuals with short sleep time ($\chi^2 = 8.59, p = 0.003$) (see Figure 9.11). Therefore, they also solved significantly more matrices ($F(1, 18) = 14.42, p = 0.001$).

We also found that during the difficult level, the participants that took sleep medication solved significantly less matrices than the participants that did not take any sleep medication ($F(1, 18) = 4.52, p = 0.047$).

Lastly, concerning the Integer Matrix task, we investigated the effect of the interaction between the sleep time and sleep quality, and the interaction between ME-type and interaction time on the task performance.

For the first interaction, we found that the quality of sleep had no impact on the performance. The performance was influenced only by the sleep time. More specifically, short sleep time leads to poor performance no matter the sleep quality. The same result was found for both difficulty levels.

For the second interaction (i.e., interaction between ME-type and task time), we did not find significant results. Nonetheless, (see Figure 9.12), evening type individuals are faster during the medium difficulty level in the afternoon session than

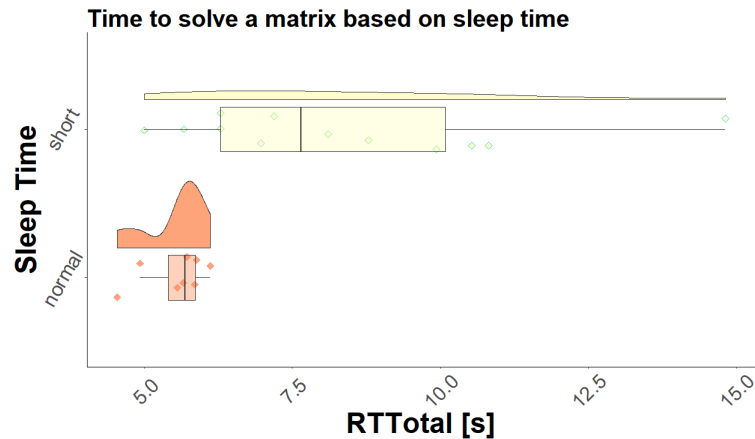


FIGURE 9.11: Time to solve a matrix based on sleep time

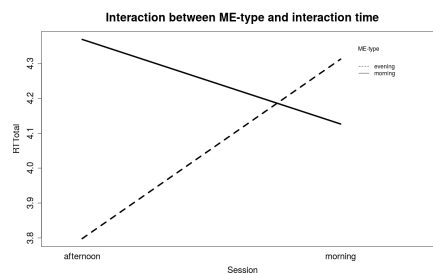


FIGURE 9.12: Interaction between ME-type and the interaction time on the average time to solve a matrix

in the morning session. Morning type individuals are faster in the morning session compared to the afternoon session. These results show a tendency, which we believe should be further investigated.

Stroop Task

For the Stroop task, we measure the task performance with two parameters: the number of correct answers and the average reaction time (RTTotal). At first, we considered the level of insomnia as factor, however we did not find any significant influence on neither of the two performance parameters. Therefore, we continued to investigate the influence of the user profile on the task performance for the participants with clinical insomnia.

None of the factors considered (i.e., age, gender, ME-type, extroversion, neuroticism) have any significant influence on the number of correct answers. For the average reaction time, we found significant results for multiple factors.

Next, we present these results.

Female participants were significantly faster than male participants ($F(1, 18) = 6.92, p = 0.017, M_{females} = 1.72, SD_{females} = 0.27, M_{males} = 2.02, SD_{males} = 0.22$). The age group of the participants is also a significant factor ($F(2, 17) = 18.58, p < 0.0001$). Pairwise comparisons using t-tests with Bonferroni correction, revealed significant differences between the middle adulthood and young/early adulthood groups ($p = 0.0089$) and between the young/early adulthood and late adulthood groups ($p < 0.0001$). Young/early adulthood individuals are significantly faster than any of the other two groups.

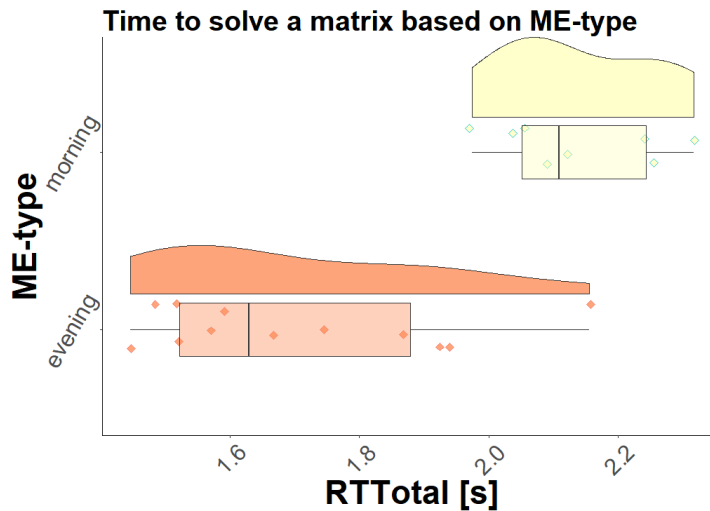


FIGURE 9.13: Average reaction time based on the ME-type

When considering the neuroticism level, we found that individuals with high neuroticism levels were significantly faster than individuals with low neuroticism levels ($F(1, 18) = 5.03, p = 0.037$). Another significant result that we found is for the ME-type. Evening type individuals are faster than morning type individuals ($F(1, 18) = 25.29, p < 0.0001$). (see Figure 9.13). No effect was found for the sleep time or sleep quality.

As with the previously presented cognitive tasks (i.e., CPT, Integer Matrix task), for this task we also investigated the interaction between the sleep time and sleep quality on the task performance, and the interaction between the ME-type and the interaction time. We did not find any significant results. However, we believe it is important should to mention them.

For the first interaction (i.e., between sleep time and sleep quality), we found that for the individuals with a medium sleep quality, the sleep time has no influence on the average time to press on the button. However, for the individuals with a bad sleep quality, a short sleep time is associated with a higher reaction time ($M = 2.01$), while a normal sleep time is associated with a shorter reaction time ($M = 1.75$). Concerning the number of correct answers, for the

For the second interaction (i.e., between ME-type and task time), we found that morning type individuals were a little bit faster during the morning session ($M = 2.11$) than during the afternoon session ($M = 2.17$). The evening type individuals were a little bit faster during the afternoon session ($M = 1.65$) than during the morning session ($M = 1.75$).

9.4.3 Physiological response analysis

For the physiological response analysis, we start with the CPT task and then continue with the cognitive games. As previously shown, the CPT task is made up of six blocks, each containing 60 trials (i.e., 20 trials for each inter-stimulus interval). If we extract the data from each subblock, as each subblock consists of a different ISI, we would not be able to compare the results. Therefore, we have decided to extract the physiological parameters from each block, and use them for analysis.

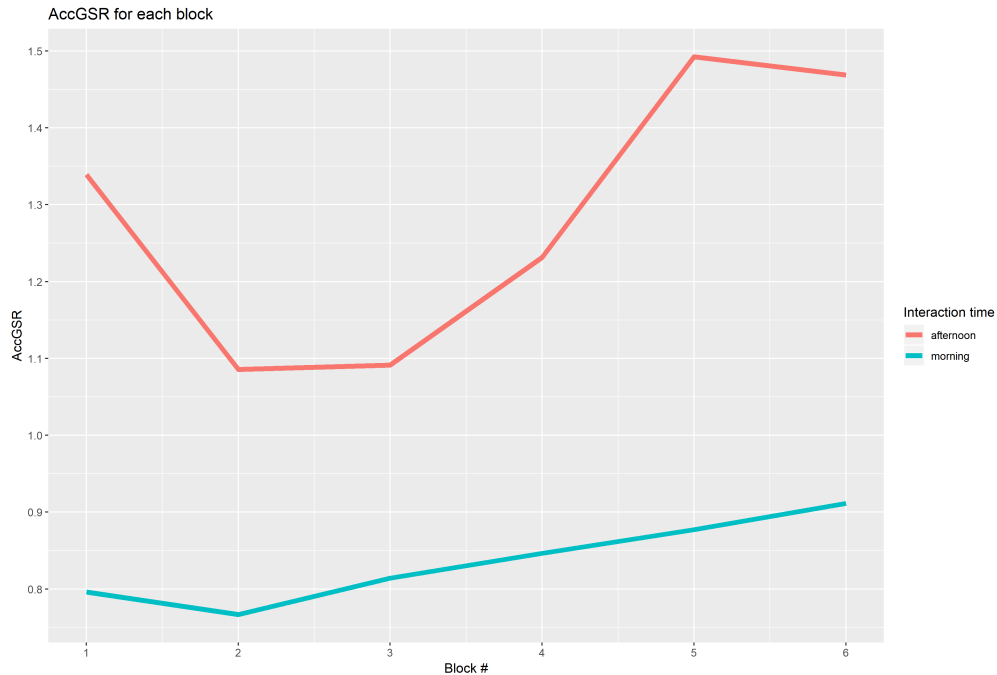


FIGURE 9.14: Variation of the AccGSR parameter over the six blocks depending on the session

CPT Task

GSR parameters

From the GSR signal we extracted three GSR parameters: the AccGSR, the number of peaks and the GSR average value. We begin the analysis with the AccGSR, we continue with the number of peaks, and finally we present the results for the average GSR value.

Our results show that in the morning individuals had lower levels for AccGSR than in the afternoon ($\chi^2 = 0.83$, $p = 0.00029$). In Figure 9.14 we show the variation of the AccGSR parameter over the six blocks for each of the two interaction sessions. In the morning, the AccGSR starts to increase from the second block onward. The increase is quite mild compare to the increase between blocks 3 and 5 for the afternoon session. Both during the morning session and during the afternoon session, the AccGSR decreases after the first block, with suggests that the participants started to get used to the task. However, as the task progressed, their arousal level started to increase again, reaching higher levels than at the beginning of the task, during both sessions. No significant results were found for the number of GSR peaks.

The third GSR parameter considered is represented by the average value for the GSR. We found that this parameter is influenced by the age group ($\chi^2 = 30.65$, $p < 0.0001$), sleep time ($\chi^2 = 8.22$, $p = 0.004$), and ME-type ($\chi^2 = 17.28$, $p < 0.0001$).

Considering the age group, we found significant differences between the middle and late adulthood age groups ($p < 0.0001$), and between the young and late adulthood age groups ($p = 0.0031$). The highest average level for the GSR was found for the late adulthood group (see also Figure 9.15 for the variation of the GSR average value over the six blocks based on the age group). After block #1 there is a slight decrease in the GSR average value, which then starts to increase again starting from block #3 (as with the AccGSR). The variation of the GSR average value is very similar for both the young and the middle adulthood groups, while for the late adulthood age

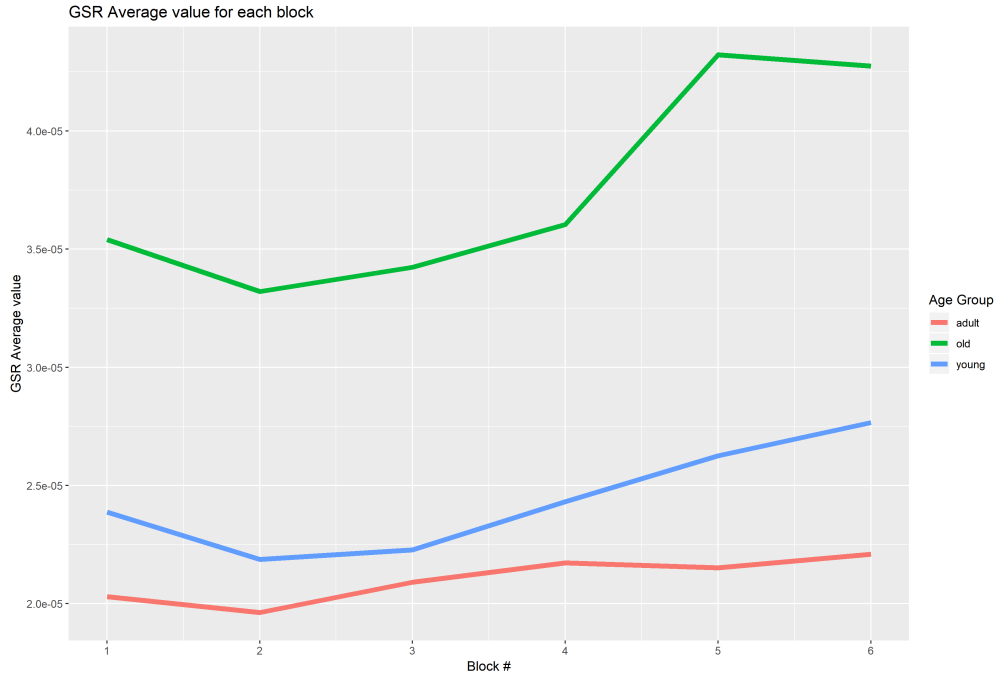


FIGURE 9.15: Variation of the GSR average value over the six blocks depending on the age group

group, the average GSR value is higher, and shows a big increase after block #4.

Concerning the sleep time, individuals with short sleep time had higher average GSR values than the individuals with normal sleep time. Based on the ME-type, we found that morning type individuals have higher average GSR level than evening type individuals.

Facial temperature

For the facial temperature variation, we consider three parameters (i.e., slope, average temperature, and range of temperatures) for each region of interest (i.e., forehead, nose, periorbital region and perinasal region).

We start with the temperature variation in the forehead region. We found that the range of temperatures in the forehead region is dependent on the neuroticism level of the participants ($\chi^2 = 8.29$, $p = 0.004$). Individuals with high neuroticism level have a lower range of temperatures than the individuals with low neuroticism levels ($M_{high N} = 1.75$, $M_{low N} = 1.94$).

For the average forehead temperature, we found that the individuals in the late adulthood age group, had significantly lower forehead average temperature than the young adulthood group ($p < 0.00001$, $M_{late adulthood} = 32.80$, $M_{young adulthood} = 33.30$), and lower forehead temperature than the middle adulthood age group ($p = 0.00014$, $M_{middle adulthood} = 33.08$). If we look how the forehead average value varies over the six blocks (see Figure 9.16), we can see that after the first block the temperature starts to decrease with almost 0.5°C and then there are only small variations between the other five blocks. The temperature shows a greater variation for the young adulthood age group than for the middle or late adulthood age group.

The forehead average temperature is also influenced by the amount of sleep the participants had the night before the task was performed ($\chi^2 = 10.62$, $p = 0.001$). Individuals with short sleep time had lower average forehead temperature than individuals with normal sleep time ($M_{short} = 32.89$, $M_{normal} = 33.21$).

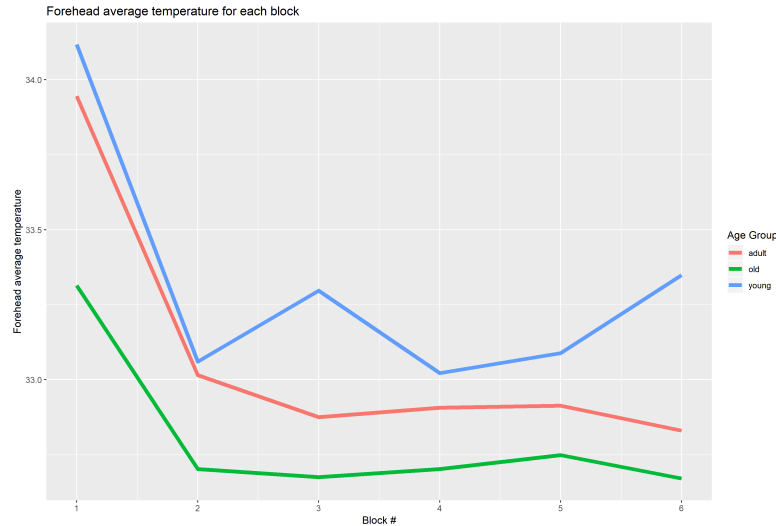


FIGURE 9.16: Forehead average temperature over the six blocks depending on the age group

Concerning the nose region, we found that the nose temperature range is higher in the afternoon session than in the morning session ($\chi^2 = 8.14$, $p = 0.004$, $M_{afternoon} = 4.16$, $M_{morning} = 3.33$). Moreover, we found that males have a higher nose average temperature ($\chi^2 = 45.26$, $p < 0.0001$, $M_{male} = 33.18$, $M_{female} = 31.41$), and that individuals with high neuroticism levels have lower average nose temperature than individuals with low neuroticism levels ($\chi^2 = 15.13$, $p = 0.0001$, $M_{high N} = 31.95$, $M_{low N} = 33.08$).

For the periorbital region, the temperature range for the participants in the middle adulthood age group was significantly lower than that of the late adulthood age group ($p = 0.0004$, $M_{middle adulthood} = 1.23$, $M_{late adulthood} = 1.65$) and that of young adulthood age group ($p = 0.0001$, $M_{young adulthood} = 1.87$). Moreover, the average temperature in the periorbital region for the participants in the young adulthood age group is significantly higher than that of the participants in the middle adulthood age group ($p < 0.0001$, $M_{young adulthood} = 33.13$, $M_{middle adulthood} = 31.70$), and than that of the participants in the late adulthood age group ($p = 0.0014$, $M_{late adulthood} = 32.63$). Sleep time is a significant factor as well ($\chi^2 = 12.79$, $p = 0.0003$), with individuals having a short sleep time having a significantly lower periorbital average temperature than the individuals with normal sleep time ($M_{normal} = 33.01$, $M_{short} = 32.60$). Moreover, individuals with bad quality sleep had lower average temperature than the individuals with medium sleep quality ($\chi^2 = 38.93$, $M_{medium} = 33.15$, $M_{bad} = 32.49$).

And lastly, for the perinasal region, males had a higher temperature range than female participants ($\chi^2 = 9.54$, $p = 0.002$, $M_{males} = 2.14$, $M_{females} = 1.62$). Moreover, male participants had a higher average temperature than female participants ($\chi^2 = 23.72$, $p < 0.0001$, $M_{males} = 33.4$, $M_{females} = 32.9$). Individuals with high neuroticism levels had lower temperature range than the individuals with high neuroticism levels ($\chi^2 = 7.8$, $p = 0.005$, $M_{high N} = 1.63$, $M_{low N} = 2.25$).

Blinking

The number of blinks was influenced by two factors: the age group ($\chi^2 = 11.98$, $p = 0.002$) and the sleep time ($\chi^2 = 8.15$, $p = 0.004$). Individuals with short sleep time blinked significantly more than the individuals with normal sleep time ($M_{normal} = 13.55$, $M_{short} = 19.91$) (see also Figure 9.17). As seen in Figure 9.18,

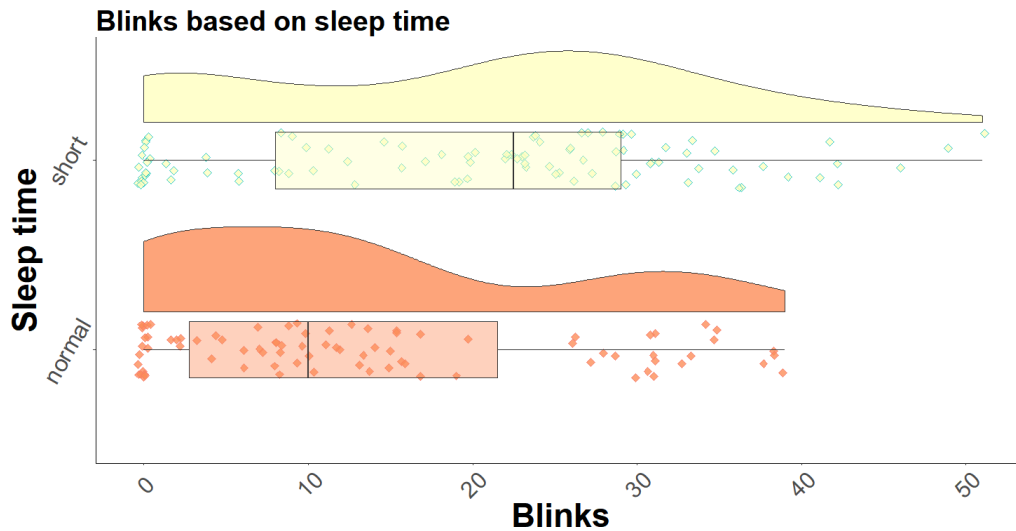


FIGURE 9.17: Blinks based on sleep time

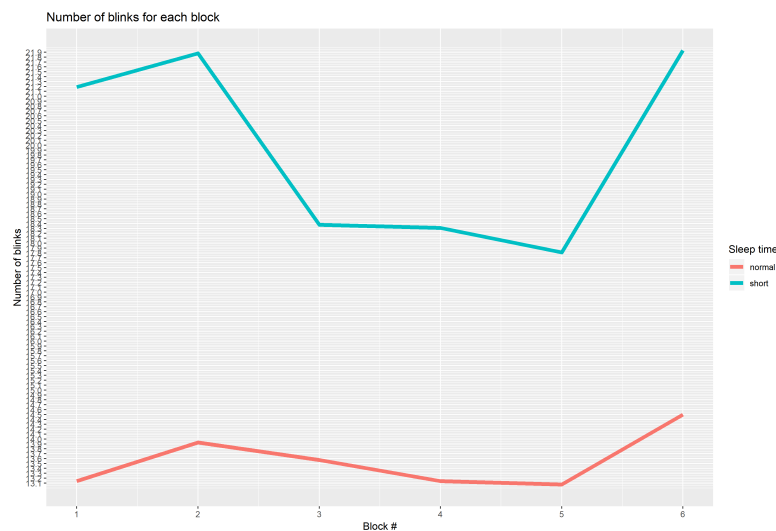


FIGURE 9.18: Blinks over the six blocks based on sleep time

the number of blinks increases after the first block, and then decreases up to block #5, after which there is a great increase again. As blinking is dependent on the task difficulty, we believe that the participants initially found the task challenging, then they got accustomed to it, while at the end, there was also the fatigue that intervened, and made the task seem more difficult. It should be mentioned again, that the CPT task lasted for a total of 14 minutes.

Individuals in the late adulthood age group blinked significantly less than the individuals in the young adulthood age group ($p = 0.015$, $M_{late\ adulthood} = 11.81$, $M_{young\ adulthood} = 17.66$) and less than the individuals in the middle adulthood age group ($p = 0.005$, $M_{middle\ adulthood} = 20.75$).

Correlation results

Lastly, we wanted to investigate the correlation between the different parameters. As the task had a total time of 14 minutes, we believed it would be more revealing to look at the correlations of the different parameters for each of the six blocks. This way, we could see their variation over time.

TABLE 9.4: Correlation results for the performance measures

Factor1	Factor 2	CPT	Block #1	Block #2	Block #3	Block #4	Block #5	Block #6
Omissions	hitRT		-0.4					
	GSR avg val	0.22		0.36				
	Periorbital range							0.48
	Perinasal range						0.43	
Commissions	hitRT	-0.24					-0.39	-0.4
	Forehead slope						-0.52	
	Forehead avg temp					-0.42		
	Forehead range			0.4		0.43		
	Periorbital slope						0.48	
	Periorbital range	0.21						
	Perinasal range	0.16						
	Blinks	-0.18						
hitRT	Forehead avg temp	-0.2		-0.48				
	Periorbital slope						-0.57	
	Periorbital avg temp	-0.28						
	Periorbital range				0.38			

First, we looked at the overall correlations between the performance measures and the physiological parameters.

The number of omissions measure is correlated over the entire task only with the GSR average value. As the number of omissions increases, so does the average GSR value. Other physiological parameters to which the number of omissions is correlated are: the hitRT (only during Block #1), the periorbital temperature range (only during Block #6), and the temperature range in the perinasal region (only during Block #4). Only the correlation with the hitRT is negative, all other correlations are positive.

For the number of commissions we found a negative correlation over the entire task with the hitRT and with the number of blinks, and a positive correlation with the temperature range in the periorbital and perinasal regions. Almost all correlations occur in Block #4 and Block #6.

For the hitRT measure, over the entire task, we found negative correlations with the average temperature in the forehead and periorbital regions. The only positive correlation was found for the periorbital temperature range, but only during Block #3.

Next, we look at the correlations over the six blocks between the physiological parameters.

In Table 9.5 we show the significant correlations (i.e., with p-values less than 0.05) for the GSR parameters with the other physiological parameters for the overall task, as well as for each of the six blocks. For the AccGSR, almost all correlations are positive, except for the periorbital slope during Block#2, when the correlation is negative. The only other physiological parameter to which the AccGSR is positively correlated both over the entire task as well as during all six blocks is the average GSR value.

For the number of GSR peaks, correlations were found with the temperatures in the nose, periorbital and perinasal regions. For the periorbital region there are only negative correlations, while for the other regions there are both positive and negative correlations. The two negative correlations occur only in Block #3. Except the

TABLE 9.5: Correlation results for the GSR parameters with the temperature and blink parameters

Factor1	Factor 2	CPT	Block #1	Block #2	Block #3	Block #4	Block #5	Block #6
AccGSR	GSR peaks	0.45	0.61	0.44			0.89	0.5
	GSR avg val	0.55	0.36	0.57	0.62	0.55	0.63	0.6
	Forehead slope							0.48
	Forehead range					0.44		
	Periorbital slope		-0.48					
	Blinks	0.16				0.4		
Factor1	Factor 2	CPT	Block #1	Block #2	Block #3	Block #4	Block #5	Block #6
GSR Peaks	GSR avg val	0.23	0.58	0.58			0.6	
	Nose slope				-0.46			
	Nose avg temp			0.44				
	Periorbital range	-0.21						
	Periorbital avg temp	-0.18						
	Perinasal slope					-0.56		
	Perinasal avg temp	0.21		0.64				

correlation with the average GSR value from Block #5, no correlations were found after Block #3.

In Table 9.6 we show the significant correlations for the forehead region parameters. We can see in the Table, that correlations are found during all blocks, and during the entire task as well. There are both positive as well as negative correlations, with no clear trend over time. The forehead average temperature is positively correlated with the perinasal and with the periorbital average temperature over almost all blocks, as well as during the entire task. Moreover, the forehead temperature range is positively correlated with the perinasal, nose, and periorbital temperature range during almost all blocks and during the entire task as well. The blinks are not at all correlated with any of the forehead region parameters.

The significant correlations that were found for the nose region parameters are presented in Table 9.7. All correlations with the nose, periorbital and perinasal region parameters are positive for all blocks and over the entire task as well. The only negative correlation was found for nose temperature range with the number of blinks (during the entire task and during Block #1).

As with the nose region parameters, for the periorbital region parameters (see Table 9.8) the only negative correlation was found for the average temperature with the number of blinks. All other correlations with the periorbital and perinasal region parameters are positive. Correlations were found during all blocks and during the entire task as well.

In Table 9.9 we present the significant correlations that were found for the perinasal region parameters. Only one correlation was found over the entire task (i.e., the negative correlation between the temperature range and the number of blinks). One interesting result is represented by the correlation between the slope of the perinasal temperature and the average temperature in the region. During Block #2 and Block #4, the correlation is negative, while in Block #3 and Block #6, the correlation is positive.

TABLE 9.6: Correlation results for the forehead region parameters with the other physiological parameters

Factor1	Factor 2	CPT	Block #1	Block #2	Block #3	Block #4	Block #5	Block #6
Forehead slope	Forehead range		0.61		-0.45			
	Forehead avg temp				0.51	-0.59		
	Nose slope				0.67			-0.7
	Nose avg temp	-0.48						
	Periorbital slope	0.58	0.71	0.65	0.74	0.73		
	Periorbital range		0.73					
	Periorbital avg temp				0.46			
Perinasal slope	0.6			0.7	0.89			
Perinasal range		0.47	0.79					
Factor1	Factor 2	CPT	Block #1	Block #2	Block #3	Block #4	Block #5	Block #6
Forehead avg temp	Forehead range	-0.28						
	Nose slope				0.66			
	Nose range	-0.23						
	Nose avg temp	0.34	0.47	0.45				0.41
	Periorbital slope				0.57			
	Periorbital range	-0.23						
	Periorbital avg temp	0.5		0.39	0.46	0.53	0.38	0.59
Perinasal slope				0.7	-0.47		0.46	
Perinasal range	-0.28				-0.5			
Perinasal avg temp	0.61	0.66	0.58		0.74	0.62	0.73	
Factor1	Factor 2	CPT	Block #1	Block #2	Block #3	Block #4	Block #5	Block #6
Forehead range	Nose slope	0.24	0.42				0.51	
	Nose range	0.32	0.61		0.59			
	Periorbital slope		0.48	-0.41				
	Periorbital range	0.49	0.66	0.59	0.5	0.61	0.69	
	Periorbital avg temp		0.45					
Perinasal range	0.54	0.78	0.51	0.53	0.61	0.47	0.73	

TABLE 9.7: Correlation results for the nose region parameters with the other physiological parameters

Factor1	Factor 2	CPT	Block #1	Block #2	Block #3	Block #4	Block #5	Block #6
Nose slope	Nose range	0.26	0.43		0.67			
	Periorbital slope	0.49			0.83			
	Periorbital range	0.19						0.5
	Periorbital avg temp	0.18			0.58			
	Perinasal slope	0.82	0.76	0.64	0.91	0.52	0.87	0.86
	Perinasal avg temp				0.56			
Factor1	Factor 2	CPT	Block #1	Block #2	Block #3	Block #4	Block #5	Block #6
Nose avg temp	Periorbital avg temp	0.43		0.46	0.43	0.45	0.48	0.49
	Perinasal range	0.22		0.45				
	Perinasal avg temp	0.77	0.77	0.84	0.74	0.76	0.77	0.77
Factor1	Factor 2	CPT	Block #1	Block #2	Block #3	Block #4	Block #5	Block #6
Nose range	Periorbital slope		0.54					
	Periorbital range	0.44	0.44		0.58	0.43	0.46	0.44
	Perinasal slope		0.48					
	Perinasal range	0.59	0.75		0.71	0.41	0.67	0.61
	Blinks	-0.19	-0.5					

TABLE 9.8: Correlation results for the periorbital region parameters with the other physiological parameters

Factor1	Factor 2	CPT	Block #1	Block #2	Block #3	Block #4	Block #5	Block #6
Periorbital slope	Periorbital slope				0.83			
	Periorbital range	0.6						
	Periorbital avg temp				0.75			
	Perinasal slope	0.69	0.52	0.68	0.87	0.71		0.74
	Perinasal avg temp				0.43			
Factor1	Factor 2	CPT	Block #1	Block #2	Block #3	Block #4	Block #5	Block #6
Periorbital avg temp	Periorbital range	0.17		0.42				0.31
	Perinasal range	0.2		0.41		0.39		
	Perinasal slope				0.51			0.47
	Perinasal avg temp	0.39			0.48			0.59
	Blinks	-0.25						
								-0.41
Factor1	Factor 2	CPT	Block #1	Block #2	Block #3	Block #4	Block #5	Block #6
Periorbital range	Perinasal slope		0.66					
	Perinasal range	0.69	0.49	0.71	0.7	0.74	0.73	0.75

TABLE 9.9: Correlation results for the perinasal region parameters with the other physiological parameters

Factor1	Factor 2	CPT	Block #1	Block #2	Block #3	Block #4	Block #5	Block #6
Perinasal slope	Perinasal range		0.68					
	Perinasal avg temp			-0.48	0.48	-0.49		0.45
Factor1	Factor 2	CPT	Block #1	Block #2	Block #3	Block #4	Block #5	Block #6
Perinasal range	Blinks	-0.21						

Integer Matrix Task

GSR parameters

The AccGSR parameter is influenced by the interaction session only, for the difficult level of the task, ($\chi^2 = 4.98$, $p = 0.02$), with higher AccGSR values during the afternoon session than the morning session.

No significant results were found for the number of GSR peaks during the medium difficulty level. While, for the difficult level, we found that the number of GSR peaks is only influenced by the gender of the participants ($\chi^2 = 4.67$, $p = 0.03$), male participants having more GSR peaks during this task than female participants.

For the third GSR parameter, the average GSR value, we found that during the medium level, the significant factors is the sleep time ($\chi^2 = 7.07$, $p = 0.007$). Individuals who had a short sleep time had a higher average GSR value than the individuals who had a normal sleep time during the night before the study. The same result was found for the difficult level as well. Individuals with a short sleep time had a higher GSR average value than the individuals with a normal sleep time ($\chi^2 = 3.98$, $p = 0.045$).

During the difficult level, males showed a higher GSR average value than females ($\chi^2 = 4.13$, $p = 0.04$). Extroverted individuals had higher GSR average value than introverted individuals ($\chi^2 = 5.17$, $p = 0.02$).

Facial temperature

During the medium level: males had a higher forehead temperature range than females ($\chi^2 = 7.73$, $p = 0.005$, $M_{males} = 1.55$, $M_{females} = 0.68$). On the other hand, for the nose region, a significant result was found during the difficult level of this game, with males having a higher average temperature than female participants ($\chi^2 = 4.6$, $p = 0.03$, $M_{males} = 33.37$, $M_{females} = 31.72$).

Age influences the average temperature in the forehead region ($\chi^2 = 8.06$, $p = 0.01$). Late adulthood individuals had a higher forehead average temperature than young adulthood individuals ($p = 0.029$, $M_{late\ adulthood} = 33.7$, $M_{young\ adulthood} = 34.4$).

Evening type individuals have a higher forehead average temperature than morning type individuals ($\chi^2 = 9.85$, $p = 0.001$, $M_{morning} = 33.73$, $M_{evening} = 34.39$), but only during the medium level (see Figure 9.19). However, for the periorbital region, a significant difference was found during the difficult level (see Figure 9.20). Evening type individuals have a higher average temperature in the periorbital region than morning type individuals ($F(1, 24) = 11.82$, $p = 0.002$, $M_{morning} = 32.94$, $M_{evening} = 33.91$).

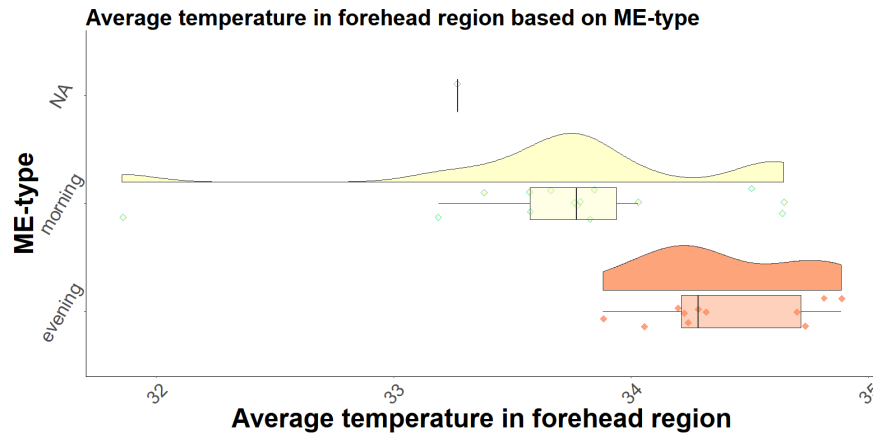


FIGURE 9.19: Average temperature in forehead region based on ME-type during Medium Integer Matrix Task

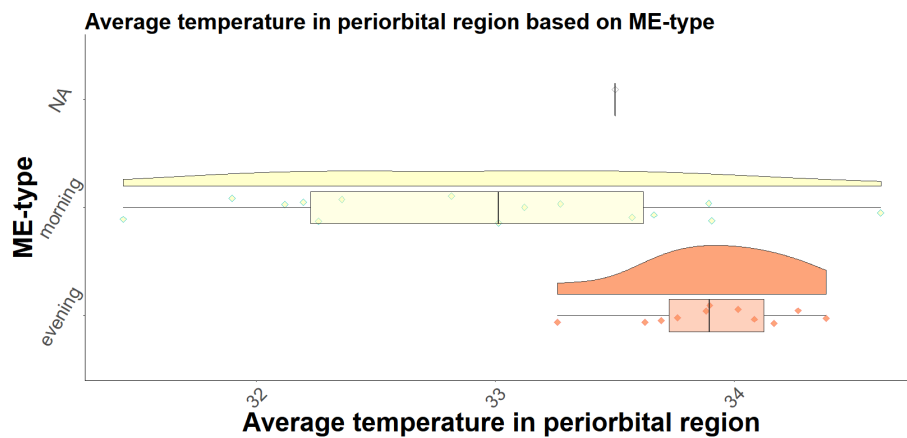


FIGURE 9.20: Average temperature in periorbital region based on ME-type during Difficult Integer Matrix Task

During the difficult level, introverted individuals had a higher forehead temperature range than extroverted individuals ($\chi^2 = 8.0$, $p = 0.004$, $M_{introverted} = 2.01$, $M_{extroverted} = 0.70$).

For the periorbital region, during the medium level: the slope for the individuals with short sleep time is positive, while for the individuals with normal sleep time it is negative ($\chi^2 = 5.0$, $p = 0.02$). Evening type individuals have a higher periorbital average temperature than morning type individuals ($\chi^2 = 11.55$, $p = 0.0006$, $M_{morning} = 32.87$, $M_{evening} = 34.06$).

Correlation results

In Table 9.10 we present the significant correlations that were found for the performance measures, as well as for the physiological parameters, during both levels of this game.

For the performance measures, we found correlations mostly for the difficult level. The only correlation found for the medium level, is that of the average time to solve a matrix and the number of correct answers, which is a negative correlation. As the time needed to solve a matrix increases, it is normal that the number of possible correct answers given decreases. The participants had a fixed amount of time to play this game. The temperature range in the perinasal region is positively correlated with the RTTotal measure and negatively correlated with the number of correct answers.

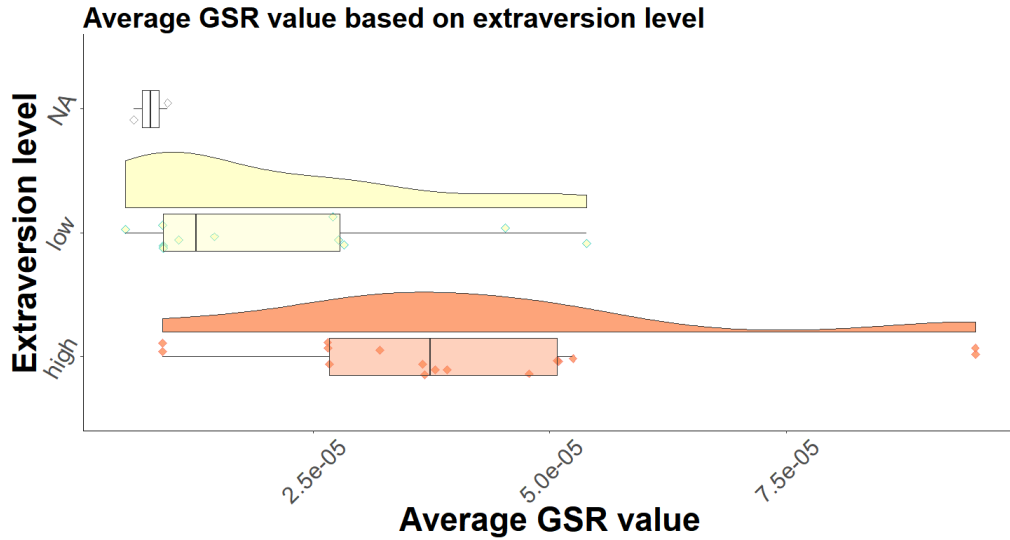


FIGURE 9.21: Average GSR value based on the extraversion level during the Stroop Task

For the physiological parameters, the majority of the correlations were found for the medium level. We found both positive correlations, as well as negative correlations between the different physiological parameters. From the correlations that were found for both difficulty levels, there are only three which change sign between the difficulty levels. For all three correlations, during the medium level there is a negative relationship, while during the difficult level, the correlation changes to a positive one. The three correlations are that between the forehead slope and the forehead average temperature, the correlation between the forehead average temperature and the temperature slope in the perinasal region, and the correlation between the periorbital average temperature and the temperature slope in the perinasal region.

Stroop Task

GSR parameters

For the GSR average value we found that the individuals with short sleep time had a higher average GSR value than the individuals with normal sleep time ($\chi^2 = 7.3$, $p = 0.006$). Extroverted individuals had higher GSR average values than introverted individuals ($\chi^2 = 4.36$, $p = 0.03$). This result is also shown in Figure 9.21.

No significant results were found for the AccGSR or the number of GSR peaks.

Facial temperature

For the nose region, individuals with high neuroticism levels had a higher temperature range than the individuals with low neuroticism level ($\chi^2 = 7.51$, $p = 0.006$, $M_{high N} = 3.49$, $M_{low N} = 1.74$). On the other hand, individuals with high neuroticism level had lower average temperature than low neuroticism level individuals ($\chi^2 = 6.4$, $p = 0.01$, $M_{high N} = 31.86$, $M_{low N} = 33.73$).

Males had a higher average temperature in the nose region than females ($\chi^2 = 6.68$, $p = 0.009$, $M_{males} = 33.48$, $M_{females} = 31.42$).

For the periorbital region, evening type individuals had a higher average temperature than morning type individuals ($\chi^2 = 8.9$, $p = 0.002$, $M_{morning} = 32.91$, $M_{evening} = 33.89$).

No significant results were found for the forehead or perinasal regions.

TABLE 9.10: Correlation results for the medium and difficult level of the Integer Matrix Task

Factor1	Factor 2	Matrix Medium	Matrix Difficult
RTTotal	Correct answers	-0.98	-0.91
	AccGSR		0.45
	Perinasal range		0.4
Correct answers	Perinasal range		-0.43
AccGSR	GSR peaks	0.59	0.66
	GSR avg val		0.42
GSR avg val	Forehead range		-0.39
Forehead slope	Forehead avg temp	-0.62	0.55
	Nose slope	0.66	
	Periorbital slope	0.7	0.73
	Periorbital avg temp	-0.66	
	Perinasal slope	0.93	0.75
	Perinasal avg temp	-0.61	
Forehead avg temp	Forehead range	-0.56	
	Nose slope	-0.53	
	Nose avg temp	0.45	0.43
	Periorbital slope	-0.64	
	Periorbital avg temp	0.57	0.43
	Perinasal slope	-0.71	0.52
	Perinasal avg temp	0.76	0.67
Forehead range	Perinasal range	-0.58	
	Nose avg temp		-0.46
	Nose range	0.49	0.68
	Periorbital slope	0.51	
	Periorbital range	0.75	0.77
	Perinasal avg temp	-0.42	-0.4
Nose slope	Perinasal range	0.83	0.64
	Nose avg temp	-0.51	
	Perinasal slope	0.96	0.62
Nose avg temp	Perinasal avg temp	-0.51	
	Nose range	-0.55	-0.55
	Periorbital avg temp	0.41	0.34
	Perinasal avg temp	0.69	0.71
Nose range	Perinsal range		-0.39
	Periorbital avg temp	-0.44	
	Periorbital range	0.73	0.61
	Perinasal avg temp	-0.42	
Periorbital slope	Perinasal range	0.68	0.82
	Periorbital avg temp	-0.71	
	Perinasal slope	0.61	0.52
	Perinasal avg temp	-0.74	
Periorbital avg temp	Perinasal range	0.44	
	Perinasal slope	-0.49	0.52
	Perinasal avg temp	0.6	0.57
Periorbital range	Perinasal range	-0.45	
	Perinasal avg val		-0.47
Perinasal slope	Perinasal range	0.77	0.74
	Perinasal avg temp	-0.61	
Perinasal avg temp	Perinasal range	-0.59	-0.44

TABLE 9.11: Correlation results for the Stroop Task

Factor1	Factor 2	Stroop
RTTotal	Correct answers	-0.47
	No answers	0.65
	Forehead avg temp	-0.51
Correct	No answers	-0.8
	Forehead avg temp	0.45
No Answer	Forehead avg temp	-0.38
AccGSR	GSR avg val	0.59
	Nose range	-0.45
	Periorbital range	-0.46
	Perinasal slope	0.55
	Perinasal range	-0.44
GSR peaks	Forehead slope	0.64
	Perinasal slope	0.55
GSR avg val	Perinasal slope	0.51
Forehead slope	Periorbital avg temp	-0.51
	Perinasal slope	0.56
Forehead avg temp	Periorbital slope	0.52
	Perinasal avg temp	0.4
Forehead range	Nose range	0.61
	Periorbital range	0.73
	Perinasal range	0.69
Nose slope	Perinasal slope	0.66
Nose avg temp	Perinasal avg temp	0.65
Nose range	Periorbital avg temp	-0.44
	Periorbital range	0.84
	Perinasal range	0.78
Periorbital avg temp	Periorbital range	-0.42
Periorbital range	Perinasal range	0.86

Correlation results

In Table 9.11 we present the significant correlations that we found.

For the performance measures, we found that all three performance measures (i.e., RTTotal, correct answers, and the number of trials without any answer) are correlated with the average temperature in the forehead region. More specifically, RTTotal and the number of trials without an answer is negatively correlated with the average temperature in the forehead region, while the number of correct answers is positively correlated.

For the physiological parameters, we found both positive correlations, as well as negative ones.

9.5 Discussion

Next, we are going to provide a short discussion based on the results presented in the previous section. For this study, we focused on two main research questions. First of all, we wanted to find out if cognitive performance is influenced by the presence or absence of insomnia. And secondly, we wanted to investigate if the user profile (i.e., age, gender, ME-type, extroversion, neuroticism) has an influence on the task performance. We considered for analysis three cognitive tasks: CPT, integer matrix task, and Stroop task.

Concerning our first research question, we found evidence that individuals with insomnia have a poorer cognitive performance than individuals without insomnia. For the CPT task the differences between the two groups are significant (i.e., for the hitRT - $p < 0.0001$, for the number of omissions - $p = 0.02$, and for the number of commissions - $p = 0.0008$). For neither the Integer Matrix task, nor the Stroop task the differences between the two groups are not significant. Therefore, we can conclude that insomnia has an impact on the task performance, but it is also dependent on the task performed.

For our second research question, we found that females have a better performance than males. Our female participants were faster during the two cognitive games, and they also committed less omissions during the CPT task. Another significant factor during all three cognitive tasks is the neuroticism personality trait. For all three tasks, individuals with high scores were significantly faster than individuals with low scores. Moreover, they also committed more omissions, but they had fewer commissions than individuals with low scores. Neuroticism is viewed as a measure of emotionality (Eysenck, 1983). High neuroticism levels are associated with restlessness, anxiousness, while low neuroticism levels are associated with being more controlled, calmer. Therefore, we could explain the faster reaction of the individuals with high neuroticism levels by considering that they are more restless and more eager to perform a given task.

For the ME-type, our results show that for the CPT and for the Stroop task, the evening type individuals are faster than morning type individuals. Evening type individuals commit more omissions in the afternoon, but they are generally faster in the afternoon than in the morning hours. For the morning type individuals we found that they are generally faster in the morning than in the afternoon hours. The majority of our participants had scores between 40 and 60, meaning that they should be in the intermediate group. However, due to the small number of participants, we decided to split our participants into only two groups (i.e., morning and evening). A new study with participants from all three groups should be performed in order to confirm our results.

Age is another significant factor for task performance. We found that for the CPT and for the Stroop task, young/early adulthood individuals are faster than the other groups of individuals. This result is in accordance with the results in (Conners et al., 2000). However, they also tend to make more mistakes. This could imply that individuals in the young/early adulthood age group are more prone to carelessness/impulsivity than individuals in the middle adulthood or late adulthood. Lastly, we found evidence that there is an interaction between the sleep quality and the sleep time during the night before the experimental task. Our significant interaction from the CPT task shows that if the sleep quality is poor, the sleep time does not influence the task performance. While, if the sleep quality is medium, a short sleep time leads to a poorer performance than a normal sleep time.

The results for the physiological response analysis show that the main factors that influence the variation of the physiological parameters during the cognitive tasks are the sleep time, the gender, and the neuroticism level of the participants. During all tasks, short sleep time was associated with higher average GSR values. As the GSR is a good indicator for the arousal level, this result suggests, that individuals who had a normal sleep time the night before the study, were more relaxed during the cognitive tasks. Sleep time was also associated with a lower average temperature in both the forehead and the periorbital regions, but only for the CPT task. It is known from the study presented in (Abdelrahman et al., 2017) that higher difficulty level is associated with an increase in forehead temperature. However, in this situation, we are not sure

how to interpret these results. Another aspect to be considered as well is that the number of participants in this study is quite low (i.e., 15 participants), therefore, more studies are needed in order to better understand the relationship between the sleep time and the variation of the facial temperature during the CPT task.

During all tasks, male participants (N=9) had higher average temperature in the nose region (all three tasks), a higher average temperature in the perinasal region (CPT task), and a higher average GSR value during the difficult level of the Integer Matrix Task. In (Christensen et al., 2012), the authors have found that male participants had significantly higher facial skin temperature than females. Therefore, our results are in accordance with the literature. Our male participants had a significantly higher facial temperature than our female participants.

During the CPT task, we found that individuals with high neuroticism level, show a lower temperature range in the forehead, nose and perinasal regions. No study was found that investigates the relationship between the facial temperature variation and the neuroticism level. Therefore, we believe further investigation is needed in order to better understand these results and to determine how they should be used in HCI and HRI scenarios.

9.6 Conclusions

In conclusion, in this Chapter we have presented our results from a preliminary study carried out with 15 participants, outpatients of a Sleep Disorder Unit. We investigated the effect of insomnia and of the user profile (i.e., age, gender, ME-type, extroversion, neuroticism) on the cognitive performance. Our results show that individuals with insomnia have a poorer performance than individuals without insomnia. Furthermore, we found evidence that the user profile influences the task performance. The user profile factors that have an effect during all three cognitive tasks are: gender and neuroticism level. Therefore, we can conclude that we found evidence to support both our research questions.

9.7 Contribution

The contribution of this chapter consisted in the investigation of the task performance in a clinical setting by using a clinically approved cognitive task. Furthermore, the methodology developed for the physiological parameters extraction and analysis during the CPT task was applied also here and we have found correlations between task performance measures and physiological parameters.

Chapter 10

Conclusion

10.1	General considerations	159
10.2	Thesis summary	159
10.2.1	Chapter 2: Related work	159
10.2.2	Chapter 3: Experimental Platforms	159
10.2.3	Chapter 4: Methodology	160
10.2.4	Chapter 5: RQ1 Relationship between cognitive performance and physiological response	160
10.2.5	Chapter 6: RQ2 Relationship between ME-type and time of the day in relation to cognitive performance	161
10.2.6	Chapter 7: RQ3 Influence of empathy, emotional intelligence and fight/flight system in HRI	162
10.2.7	Chapter 8: Assistive applications (I)	162
10.2.8	Chapter 9: Assistive applications (II)	163
10.3	Perspectives	163

10.1 General considerations

The main purpose of this thesis was to explore the relationship that exists between the morningness-eveningness type (ME-type) of an individual, the cognitive performance, and the internal physiological state of that individuals in different human-robot interaction scenarios. We have explored this relationship in laboratory settings, by using online platforms, in a real-world environment represented by the homes of elderly with mild cognitive impairment (MCI), and a clinical environment represented by the Sleep Unit of a hospital in Paris.

10.2 Thesis summary

10.2.1 Chapter 2: Related work

In this chapter, we have presented a literature review of the three main research directions related to the thesis. The research directions are: using the physiological activity of users in human-robot interaction scenarios, relationship between morningness-eveningness type (ME-type) and cognitive performance, and the physiological measures associated to cognitive load (i.e., blinking, galvanic skin response and facial temperature variation).

10.2.2 Chapter 3: Experimental Platforms

The purpose of this chapter was to present the different experimental platforms used in this thesis. We first presented the two robotic platforms used for the experiments

presented in this thesis. The two robots are Tiago and Pepper. Next, we presented the sensors used to measure the physiological parameters. More specifically, we presented the Optris thermal camera used to extract the facial temperature, the ASUS RGB-D camera to extract the blinks, and the Grove GSR sensor used to measure the galvanic skin response. Lastly, we presented the three cognitive games that we developed and used during this thesis: the Stroop Game, the Integer Matrix Task, and the Decimal Matrix Task.

10.2.3 Chapter 4: Methodology

Chapter 4 was divided into two main parts: the presentation of the methodology for the extraction and analysis of the physiological parameters, and the presentation of the psychological questionnaires used to determine the user profile. The physiological parameters that we decided to extract and analyse are: the facial temperature variation, the blinking, and the GSR. The psychological questionnaires presented are the Morningness-Eveningness Questionnaire, the Reinforcement Sensitivity Theory Personality Questionnaire, the Eysenck Personality Questionnaire, and the Adult/Adolescent Sensory Profile Questionnaire.

The main contributions of this chapter consist in the:

1. Training of a model to detect faces in thermal images
2. Training of a model to predict 11 facial feature points in thermal images
3. Definition of eight regions of interest for the face from which multiple parameters can be extracted and analysed
4. Development of a method to detect blinks both on-line, as well as, off-line
5. Development of a software to extract and analyse multiple GSR features

10.2.4 Chapter 5: RQ1 Relationship between cognitive performance and physiological response

The purpose of Chapter 5 was to investigate the relationship between cognitive performance and the associated physiological response. A within participants study was designed around three conditions: an interaction without robot, an interaction with an encouraging robot, and an interaction with a stressing robot. The robotic platform used for this experiment was Tiago. The experimental task consisted of the non-verbal word-color Stroop Task. The physiological sensors investigated are the GSR, the blinks, and the facial temperature variation.

Our results showed that there is a slight interaction between the time of the day when a task is performed, the ME-type and the cognitive performance. We also found that the most important physiological parameters that enable us to distinguish between the two robot conditions are the AccGSR and the temperature variation in the forehead region (we obtained a classification accuracy of 79.2%).

The main contributions of this chapter consist in:

1. Using the MEQ and the sensory profile in HRI
2. Finding the interaction between the ME-type and the time of the day when the task was performed (we had only two interaction times: morning - before 15h00, and afternoon - after 15h00)
3. Finding a relationship between the ME-type and the sensory profile

Some of the limitations of the study presented in this chapter consist in:

- A relative small number of participants (i.e., N=24)
- Only two interaction times were considered (i.e., morning - before 15h00, and afternoon - after 15h00)
- The encouraging/stressing content provided by the robot did not take into consideration the task performance, we believe this might have influenced how the participants reacted during the task.

Some of the perspectives based on this study are:

- To design a study in which the robot provides the encouraging/stressing content only after the user provides an answer, and one which is in accordance with the response
- To further investigate the relationship between the auditory and visual stimuli that can be used by a robot and the sensory profile of the individual it interacts with
- To further investigate the relationship between ME-type, time of the day and cognitive performance

10.2.5 Chapter 6: RQ2 Relationship between ME-type and time of the day in relation to cognitive performance

Based on the results from Chapter 5, we have decided to further investigate the relationship between ME-type, the time of the day when the cognitive task is performed, and the task performance. For this purpose, an online study was designed and 135 participants were recruited to take part in it. The cognitive tasks could be performed whenever and wherever the users wanted. As in our previous study we only considered two interaction times, for this study, we considered four main times of the day: morning (between 06h00 and 12h59), afternoon (between 13h00 and 16h59), evening (between 17h00 and 20h59) and night (between 21h00 and 05h59). The experimental task consisted in the Stroop Game, the Integer Matrix Task and the Decimal Matrix Task. We considered for the analysis the two best performances of each participant.

Our results showed that the performance is dependent on the game, the time of the day when the game is played, and the user profile of that individual. We also found that the influence of the interaction between time of the day and task performance is dependent on the task performed. The strongest influence was found for the Stroop game, while the weakest one for the Decimal Matrix Task. Our main take away from this study is that there is still no clear time of the day when a specific type of individual has a better or worse performance; further studies are required to better understand this relationship and to determine how this can be used in social robotics.

The main contributions of this chapter consist in:

1. We developed an online platform where participants could perform the experimental task
2. This is a large scale non controlled study
3. We investigated the performance of Morning type individuals, Evening type individuals, as well as that of Intermediate type individuals. There are few studies who also consider the intermediate type individuals

4. In this study we investigated three cognitive tasks

Some of the perspectives based on this study are:

- To recruit more participants to perform the experimental tasks
- To design a robot behavior that can propose different activities (e.g., cognitive, physical) by taking into consideration the time of the day and the ME-type of the individual

10.2.6 Chapter 7: RQ3 Influence of empathy, emotional intelligence and fight/flight system in HRI

The purpose of Chapter 7 was to understand the emotional state of an individual while interacting with a robot around a task, in an unforeseen situation. More specifically, while playing the Jenga game, the participants were abruptly interrupted by the experimenter by either knocking over their tower, or by just informing them that they should stop the task. We also investigated if the participants reacted differently when they interacted with a robot or with a human.

Our results show significant differences in the GSR parameters based on condition type (knocking over the tower, or just informing the participants that they should stop) and the experimenter. The participants that interacted with the robot, needed a longer time for their GSR signal to start to increase. Furthermore, the GSR amplitude during the condition when the tower was knocked over by the experimenter, was higher than when the participants were asked by the experimenter to stop. We also found a relationship between the facial temperature variation and the fight/flight/freeze system.

The main contributions of this chapter consist in:

1. The investigation of the emotional response during unforeseen situation
2. The usage of the Fight/Flight system in HRI. To the best of our knowledge this has not been used before in HRI
3. We found a relationship between the FFFS and the variation of the different physiological parameters

10.2.7 Chapter 8: Assistive applications (I)

In Chapter 8, we have presented how the methodology developed by us was used in a real-world environment with a vulnerable population represented by the elderly with mild cognitive impairment (MCI). We present three studies that were carried out as part of the H2020 ENRICHME research project. The first study was carried out during the first year of the research project, with the purpose of finding out what the elderly thought of the robotic system and of the applications developed for them. Based on their feedback, the graphical user interface was redesigned and more applications were included. Some of the first lessons that we learned are also presented. For the second study, we recruited one of the final testing users of the ENRICHME project for a 5-day interaction scenario with two sessions played for each day. We investigated the outcome of the intensive cognitive stimulation, as well as, how the variation of the physiological parameters varied depending on the task performed. The third study presented, represents the final testing phase of the ENRICHME project. We have presented how the 11 end users used the human-machine interface developed for them.

The main contributions of this chapter consist in:

1. The development of the ENRICHME graphical user interface and the nine applications included
2. Two short term studies were carried out with the elderly with MCI

10.2.8 Chapter 9: Assistive applications (II)

In Chapter 9, we have presented how the methodology developed by us was used in a clinical environment with a vulnerable population represented by the individuals suffering from different sleep disorders. The experiment was carried out in the Sleep Unit from the University Hospitals Pitié Salpêtrière - Charles Foix in Paris. We investigated the effect of insomnia and user profile (i.e., age, gender, ME-type, extroversion and neuroticism).

Our results confirm what was already shown in the literature, that individuals with insomnia have a poorer cognitive performance than the individuals without insomnia. The results for the the physiological response analysis show that the main factors that influence the variation of the physiological parameters during the cognitive tasks are the sleep time, the gender, and the neuroticism level of the participants.

The main contributions of this chapter consist in:

1. The investigation of cognitive performance in a clinical setting by using a clinically approved cognitive task
2. The methodology presented in Chapter 4 was used to extract and analyse the physiological parameters

10.3 Perspectives

The main purpose of this thesis was to explore the relationship between ME-type, cognitive performance, and user internal state in different HRI scenarios. For this purpose, we have developed a methodology to measure and analyse physiological parameters, we have used different psychological questionnaires to determine the user profile and we have tested our system in laboratory setting as well as in real-world environments.

This thesis was mostly focused on how to measure the physiological parameters in different human-robot interaction scenarios. While this is important, it represents only a first step in a successful interaction between an individual and a robotic platform. Our next step will consist in enabling the robotic platforms to adapt their behavior based on the internal state of the individuals they interact with. This means, that we should attach meaning to the variation of the physiological parameters investigated. What emotion do they represent? Is it universal? Does it depend on the context? We believe that once this is accomplished, social robots will have a greater presence in our everyday lives. For this purpose, first we will need to select the parameters that we want to include in our emotion model and then to create a database with the variation of these parameters depending on the known emotional state of multiple individuals. Once the database is created a learning process can be applied, so that the robot can learn to differentiate between different emotional states based on the physiological parameters.

Furthermore, another important source of information concerning the internal state of an individual is represented by the facial expressions. They can be detected by using an RGB camera, therefore no invasive sensors are required. Some solutions are already available, however, they consider only a limited number of facial expressions,

and some of them are difficult to integrate with robotic platforms. Therefore, our future work is going to be focused on enabling robotic systems to recognize facial expressions and the underlying emotions that they represent. By using both the physiological parameters, as well as the facial expressions, the robot will be able to have a better understanding of the emotional state of the individual it interacts with.

However, this is just the first step. Next, the robot needs to be able to learn how it should behave depending on the context and the emotional state of the individual it interacts with. Therefore, our work will be focused on determining what components need to be adapted (e.g., speech, gestures, movement speed, gaze, etc) and how (e.g., speaking more loudly, displaying more or less gestures, looking directly at the individual, moving faster or slower) depending on the emotional state of the individual. There are some recommendations in the literature that are going to be investigated and applied in different HRI scenarios.

Another aspect that needs to get more attention is the improvement of the methodology developed in this thesis. As an example, for the facial regions of interest, we want to take into consideration the rotation of the head as well, so that the regions can be rotated. This will ensure that important data is not lost, and that we eliminate temperatures, which are associated with other facial regions, or with the background. Moreover, the blink detection algorithm is also going to be improved. Different machine learning algorithms are going to be applied in order to have a more reliable blink detection algorithm.

As seen in Chapter 2, there are also other physiological signals that are worth investigating. Two such signals are represented by the heart rate and the respiration rate. An idea would be to find reliable non-invasive methods to measure them, methods that can be used in HRI scenarios in real-time.

Appendix A

List of Publications

Refereed Book Chapters

1. Ferland François; Agrigoroaie Roxana; Tapus Adriana, *Assistive humanoid robots for the elderly with mild cognitive impairment*, Humanoid Robotics: A reference, 2018

Refereed Journal Papers

1. Agrigoroaie Roxana; Ciocirlan Stefan Dan; Tapus Adriana, *In the Wild HRI Scenario: Influence of Regulatory Focus Theory*, Frontiers in Robotics and AI - Under Review
2. Agrigoroaie Roxana; Pallanca Olivier; Tapus Adriana, *Impact of Insomnia and User Profile on Cognitive Performance*, IEEE Transactions on Affective Computing - Under Review
3. Serhan Coşar; Manuel Fernandez-Carmona; Roxana Agrigoroaie; Jordi Pages; François Ferland; Feng Zhao; Shigang Yue; Nicola Bellotto; Adriana Tapus, *ENRICHME: Perception and Interaction of an Assistive Robot for the Elderly at Home*, International Journal of Social Robotics - Under Review
4. Agrigoroaie Roxana; Tapus Adriana, *Cognitive performance and physiological response analysis*, International Journal of Social Robotics, 2019

Refereed Conference Papers

1. Ciocirlan Stefan Dan; Agrigoroaie Roxana; Tapus Adriana, *Human-Robot Team: Effects of Communication in Analyzing Trust*, IEEE International Conference on Robot and Human Interactive Communication (ROMAN 2019) - Under Review
2. Agrigoroaie Roxana; Tapus Adriana, *Physiological differences depending on task performed in a 5-day interaction scenario designed for the elderly: A pilot study*, International Conference on Social Robotics (ICSR 2018)
3. Agrigoroaie Roxana; Cruz-Maya Arturo; Tapus Adriana, *Oh! I am so sorry!: Understanding user physiological variation while spoiling a game task*, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2018)
4. Agrigoroaie Roxana; Tapus Adriana, *Physiological parameters variation based on the sensory stimuli used by a robot in a news reading task*, IEEE International Conference on Robot and Human Interactive Communication (ROMAN 2018)

5. Agrigoroaie Roxana; Tapus Adriana, *The outcome of a week of intensive cognitive stimulation in an elderly care setup: A pilot test*, IEEE International Conference on Robot and Human Interactive Communication (ROMAN 2018), Special Session: Assistive Robotics for Elderly Care
6. Agrigoroaie Roxana; Tapus Adriana, *Influence of robot's interaction style on performance in a stroop task*, International Conference on Social Robotics (ICSR2017)
7. Agrigoroaie Roxana; Ciocirlan Stefan Dan; Tapus Adriana, *News application adaptation based on user sensory profile*, International Conference on Social Robotics (ICSR2017)
8. Cruz-Maya Arturo; Agrigoroaie Roxana; Tapus Adriana, *Improving user's performance by motivation: matching robot interaction strategy with user's regulatory state*, International Conference on Social Robotics (ICSR2017)
9. Agrigoroaie Roxana; Tapus Adriana, *Contactless physiological data analysis for user quality of life improving by using a humanoid social robot*, International Conference on Image Analysis and Processing (ICIAP 2017)
10. Agrigoroaie Roxana; Ferland François; Tapus Adriana, *The ENRICHME project: lessons learnt from a first interaction with the elderly*, International Conference on Social Robotics (ICSR2016)

Refereed Workshop and Symposia Papers

1. Agrigoroaie Roxana; Ferland François; Tapus Adriana, *The ENRICHME Project: Validation of the Human-Machine Interface*, Mobile Robot Assistans for the Elderly Workshop, part of ICRA2019
2. Agrigoroaie Roxana; Pallanca Olivier; Tapus Adriana, *Impact of insomnia and morningness-eveningness type on cognitive performance*, Journee Fedev 2018
3. Agrigoroaie Roxana; Tapus Adriana, *The influence of empath, mood and self control in a human-robot interaction scenario*, Journee Robotique et IA 2018
4. Agrigoroaie Roxana; Tapus Adriana, *Understanding how physiological data can help a humanoid robot adapt to the needs of the elderly with mild cognitive impairment*, Handiversite 2018
5. Agrigoroaie Roxana; Tapus Adriana, *Defining the user profile for the behavior adaptation of a robot*, Journee Fedev 2017
6. Agrigoroaie Roxana; Tapus Adriana, *Developing a healthcare robot with personalized behaviors and social skills for the elderly*, Pioneers Workshop, part of HRI2016

Refereed Poster Papers

1. Agrigoroaie Roxana; Tapus Adriana, *Detecting deception in a human-robot interaction scenario*, Late-Breaking Report ACM/IEEE International Conference on Human-Robot Interaction (HRI 2017)

Appendix B

Morningness Eveningness Questionnaire

For each question, please select the answer that best describes you by circling the point value that best indicates how you have felt in recent weeks.

1. *Approximately* what time would you get up if you were entirely free to plan your day?

- (5) 5:00 AM - 6:30 AM (05:00 - 06:30 h)
- (4) 6:30 AM - 7:45 AM (06:30 - 07:45 h)
- (3) 7:45 AM - 9:45 AM (07:45 - 09:45 h)
- (2) 9:45 AM - 11:00 AM (09:45 - 11:00 h)
- (1) 11:00 AM - 12 noon (11:00 - 12:00 h)

2. *Approximately* what time would you go to bed if you were entirely free to plan your evening?

- (5) 8:00 PM - 9:00 PM (20:00 - 21:00 h)
- (4) 9:00 PM - 10:15 PM (21:00 - 22:15 h)
- (3) 10:15 PM - 12:30 AM (22:15 - 00:30 h)
- (2) 12:30 AM - 1:45 AM (00:30 - 01:45 h)
- (1) 1:45 AM - 3:00 AM (01:45 - 03:00 h)

3. If you usually have to get up at a specific time in the morning, how much do you depend on the alarm clock?

- (4) Not at all (3) Slightly (2) Somewhat (1) Very much

4. How easy do you find it to get up in the morning (when you are not awakened unexpectedly)?

- (1) Very difficult (2) Somewhat difficult (3) Fairly easy (4) Very easy

5. How alert do you feel during the first half hour after you wake up in the morning?

- (1) Not at all alert (2) Slightly alert (3) Fairly alert (4) Very alert

6. How hungry do you feel during the first half hour after you wake up?

- (1) Not at all hungry (2) Slightly hungry (3) Fairly hungry (4) Very hungry

7. During the first half hour after you wake up in the morning, how do you feel?

(1) Very tired (2) Fairly tired (3) Fairly refreshed (4) Very refreshed

8. If you had no commitments the next day, what time would you go to bed compared to your usual bedtime?

(4) Seldom or never later (3) Less than 1 hour later
(2) 1-2 hours later (1) More than 2 hours later

9. You have decided to do physical exercise. A friend suggests that you do this for one hour twice a week, and the best time for him is between 7-8 AM. Bearing in mind nothing but your own '**internal clock**', how do you think you would perform?

(4) Would be in a good form (3) Would be in reasonable form
(2) Would find it difficult (1) Would find it very difficult

10. At *approximately* what time in the evening do you feel tired, and, as a result, in need of sleep?

(5) 8:00 PM - 9:00 PM (20:00 - 21:00 h)
(4) 9:00 PM - 10:15 PM (21:00 - 22:15 h)
(3) 10:15 PM - 12:45 AM (22:15 - 00:45 h)
(2) 12:45 AM - 2:00 AM (00:45 - 02:00 h)
(1) 2:00 AM - 3:00 AM (02:00 - 03:00 h)

11. You want to be at your peak performance for a test that you know is going to be mentally exhausting and will last two hours. You are entirely free to plan your day. Considering only your '**internal clock**', which one of the four testing times would you choose?

(6) 8 AM - 10 AM (4) 11 AM - 1 PM (2) 3 PM - 5 PM (0) 7 PM - 9 PM

12. If you got into bed at 11 PM, how tired would you be?

(0) Not at all tired (2) A little tired (3) Fairly tired (5) Very tired

13. For some reason you have gone to bed several hours later than usual, but there is no need to get up at any particular time the next morning. Which one of the following are you most likely to do?

(4) Will wake up at usual time, but will not fall back asleep
(3) Will wake up at usual time and will doze thereafter
(2) Will wake up at usual time, but will fall asleep again
(1) Will not wake up until later than usual

14. One night you have to remain awake between 4 - 6 AM in order to carry out a night watch. You have no time commitments the next day. Which one of the alternatives would suit you best?

(1) Would not go to bed until the watch is over
(2) Would take a nap before and sleep after
(3) Would take a good sleep before and nap after
(4) Would sleep only before the watch

15. You have two hours of hard physical work. You are entirely free to plan your day. Considering only your '**internal clock**', which of the following times would you choose?

- (4) 8 AM - 10 AM (3) 11 AM - 1 PM (2) 3 PM - 5 PM (1) 7 PM - 9 PM

16. You have decided to do physical exercise. A friend suggests that you do this for one hour twice a week. The best time for her is between 10-11 PM. Bearing in mind only your '**internal clock**', how well do you think you would perform?

- (1) Would be in good form (2) Would be in reasonable form
(3) Would find it difficult (4) Would find it very difficult

17. Suppose you can choose your own work hours. Assume that you work a five-hour day (including breaks), your job is interesting, and you are paid based on your performance. At *approximately* what time would you choose to begin?

- (5) 5 hours starting between 4 - 8 AM (05 - 08 h)
(4) 5 hours starting between 8 - 9 AM (08 - 09 h)
(3) 5 hours starting between 9 AM - 2 PM (09 - 14 h)
(2) 5 hours starting between 2 - 5 PM (14 - 17 h)
(1) 5 hours starting between 5 PM - 4 AM (17 - 04 h)

18. At *approximately* what time of day do you usually feel your best?

- (5) 5 - 8 AM (05 - 08 h) (4) 8 - 10 AM (08 - 10 h)
(3) 10 AM - 5 PM (10 - 17 h) (2) 5 - 10 PM (17 - 22 h)
(1) 10 PM - 5 AM (22 - 05 h)

19. One hears about '**morning types**' and '**evening types**'. Which one of these types do you consider yourself to be?

- (6) Definitely a morning type
(4) Rather more a morning type than an evening type
(2) Rather more an evening type than a morning type
(1) Definitely an evening type

Computing the Morningness-Eveningness score

This questionnaire has 19 questions, each with a number of points. First, add up the points you circled. Scores can range from 16 to 86 (See Table ??). Scores of 41 and below indicate "evening types". Scores of 59 and above indicate "morning types". Scores between 42-58 indicate "intermediate types".

TABLE B.1: Scores and ME-types

16-30	31-41	42-58	59-69	70-86
definite evening	moderate evening	intermediate	moderate morning	definite morning

Appendix C

Regulatory Focus Questionnaire - proverb form

For every proverb, indicate in which measure this can be applied to your way to see life in general (mark the corresponding number).

	Response						
	Not at all						Absolutely
1. Nothing ventured, nothing gained.	1	2	3	4	5	6	7
2. He is no wise man that will quit a certainty for an uncertainty.	1	2	3	4	5	6	7
3. Slow and steady wins the race.	1	2	3	4	5	6	7
4. Better an egg today than a hen tomorrow.	1	2	3	4	5	6	7
5. To seize the opportunity.	1	2	3	4	5	6	7
6. When in doubt, don't.	1	2	3	4	5	6	7
7. Too much caution does not reach its goal.	1	2	3	4	5	6	7
8. A bird in the hand is worth two in the bush.	1	2	3	4	5	6	7
9. We know what we left but we do not know what is found.	1	2	3	4	5	6	7
10. Nothing is impossible for a willing heart.	1	2	3	4	5	6	7
11. Save for a rainy day.	1	2	3	4	5	6	7
12. You have to turn the mill when the wind blows.	1	2	3	4	5	6	7
13. Prevention is better than cure.	1	2	3	4	5	6	7
14. Fortune favours the bold.	1	2	3	4	5	6	7
15. No pain, no gain.	1	2	3	4	5	6	7
16. Better safe than sorry.	1	2	3	4	5	6	7
17. Don't put all your eggs in the same basket.	1	2	3	4	5	6	7
18. Where there is a will there is a way.	1	2	3	4	5	6	7

Computing the score

Promotion focus = $(1 + 5 + 7 + 10 + 12 + 14 + 15 + 18) / 8$

Prevention focus = $(2 + 3 + 4 + 6 + 8 + 9 + 11 + 13 + 16 + 17) / 10$

Appendix D

Eysenck Personality Questionnaire

Please answer each question by selecting either YES or NO. There are no right or wrong answers and no trick questions. Work quickly and do not think too long about the exact meaning of the questions.

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

30. Would you call yourself tense or 'highly-strung'?	YES	NO
31. Do you think people spend too much time safeguarding their future with savings and insurances?	YES	NO
32. Do you like mixing with people?	YES	NO
33. As a child were you ever cheeky to your parents?	YES	NO
34. Do you worry too long after an embarrassing experience?	YES	NO
35. Do you try not to be rude to people?	YES	NO
36. Do you like plenty of bustle and excitement around you?	YES	NO
37. Have you ever cheated at a game?	YES	NO
38. Do you suffer from 'nerves'?	YES	NO
39. Would you like other people to be afraid of you?	YES	NO
40. Have you ever taken advantage of someone?	YES	NO
41. Are you mostly quiet when you are with other people?	YES	NO
42. Do you often feel lonely?	YES	NO
43. Is it better to follow society's rules than go your own way?	YES	NO
44. Do other people think of you as being very lively?	YES	NO
45. Do you always practice what you preach?	YES	NO
46. Are you often troubled about feelings of guilt?	YES	NO
47. Do you sometimes put off until tomorrow what you ought to do today?	YES	NO
48. Can you get a party going?	YES	NO

Scoring key:

Psychotism scale:

YES: 10, 14, 22, 31, 39

NO: 2, 6, 18, 26, 28, 35, 43

Extroversion scale:

YES: 3, 7, 11, 15, 19, 23, 32, 36, 44, 48

NO: 27, 41

Neuroticism scale:

YES: 1, 5, 9, 13, 17, 21, 25, 30, 34, 38, 42, 46

Lie scale:

YES: 4, 16, 45

NO: 8, 12, 20, 24, 29, 33, 37, 40, 47

A point is given for each question that has an answer that corresponds to the personality trait keys.

Appendix E

Reinforcement Sensitivity Theory Personality Questionnaire

Below are a list of statements about everyday feelings and behaviors. Please rate how accurately each statement describes *you in general*. Circle only one response. Do not spend too much time thinking about the questions and please answer honestly.

How accurately does each statement describe *you*

	Response			
	Not at all	Slightly	Mode rately	Highly
1. I feel sad when I suffer even minor setbacks.	1	2	3	4
2. I am often preoccupied with unpleasant thoughts.	1	2	3	4
3. Sometimes even little things in life can give me great pleasure.	1	2	3	4
4. I am especially sensitive to reward.	1	2	3	4
5. I put in a big effort to accomplish important goals in my life.	1	2	3	4
6. I sometimes feel 'blue' for no good reason.	1	2	3	4
7. When feeling 'down', I tend to stay away from people.	1	2	3	4
8. I often experience a surge of pleasure running through my body.	1	2	3	4
9. I would be frozen to the spot by the sight of a snake or spider.	1	2	3	4
10. I have often spent a lot of time on my own to "get away from it all"	1	2	3	4
11. I am a very active person.	1	2	3	4
12. I'm motivated to be successful in my personal life.	1	2	3	4
13. I am always 'on the go'.	1	2	3	4
14. I regularly try new activities just to see if I enjoy them.	1	2	3	4
15. I get carried away by new projects.	1	2	3	4
16. Good news makes me feel over-joyed.	1	2	3	4
17. The thought of mistakes in my work worries me.	1	2	3	4
18. When nervous, I sometimes find my thoughts are interrupted.	1	2	3	4
19. I would run quickly if fire alarms in a shopping mall started ringing.	1	2	3	4

How accurately does each statement describe <i>you</i>	Response			
	Not at all	Slightly	Mod- erately	Highly
20. I often overcome hurdles to achieve my ambitions.	1	2	3	4
21. I often feel depressed.	1	2	3	4
22. I think I should 'stop and think' more instead of jumping into things too quickly.	1	2	3	4
23. I often feel that I am on an emotional 'high'.	1	2	3	4
24. I love winning competitions.	1	2	3	4
25. I get a special thrill when I am praised for something I've done well.	1	2	3	4
26. I take a great deal of interest in hobbies.	1	2	3	4
27. I sometimes cannot stop myself talking when I know I should keep my mouth closed.	1	2	3	4
28. I often do risky things without thinking of the consequences.	1	2	3	4
29. My mind is sometimes dominated by thoughts of the bad things I've done.	1	2	3	4
30. I get very excited when I get what I want.	1	2	3	4
31. I feel driven to succeed in my chosen career.	1	2	3	4
32. I'm always finding new and interesting things to do.	1	2	3	4
33. I'm always weighing-up the risk of bad things happening in my life.	1	2	3	4
34. People are often telling me not to worry.	1	2	3	4
35. I am very open to new experiences in life.	1	2	3	4
36. I always celebrate when I accomplish something important.	1	2	3	4
37. I find myself reacting strongly to pleasurable things in life.	1	2	3	4
38. I find myself doing things on the spur of the moment.	1	2	3	4
39. I would instantly freeze if I opened the door to find a stranger in the house.	1	2	3	4
40. I'm always buying things on impulse.	1	2	3	4
41. I am very persistent in achieving my goals.	1	2	3	4
42. When trying to make a decision, I find myself constantly chewing it over.	1	2	3	4
43. I often worry about letting down other people.	1	2	3	4
44. I would go on a holiday at the last minute.	1	2	3	4
45. I would run fast if I knew someone was following me late at night.	1	2	3	4
46. I would leave the park if I saw a group of dogs running around barking at people.	1	2	3	4
47. I worry a lot.	1	2	3	4
48. I would freeze if I was on a turbulent aircraft.	1	2	3	4
49. My behavior is easily interrupted.	1	2	3	4
50. It's difficult to get some things out of my mind.	1	2	3	4
51. I think the best nights out are unplanned.	1	2	3	4
52. There are some things that I simply cannot go near.	1	2	3	4
53. If I see something I want, I act straight away.	1	2	3	4

How accurately does each statement describe <i>you</i>	Response			
	Not at all	Slightly	Mod- erately	Highly
54. I think it is necessary to make plans in order to get what you want in life.	1	2	3	4
55. When nervous, I find it hard to say the right words.	1	2	3	4
56. I find myself thinking about the same thing over and over again.	1	2	3	4
57. I often wake up with many thoughts running through my mind.	1	2	3	4
58. I would not hold a snake or spider.	1	2	3	4
59. Looking down from a great height makes me freeze.	1	2	3	4
60. I often find myself 'going into my shell'.	1	2	3	4
61. My mind is dominated by recurring thoughts.	1	2	3	4
62. I am the sort of person who easily freezes-up when scared.	1	2	3	4
63. I take a long time to make decisions.	1	2	3	4
64. I often find myself lost for words.	1	2	3	4
65. I will actively put plans in place to accomplish goals in my life.	1	2	3	4

Scoring key

Fight-Flight-Freeze System (FFFS): 9, 19, 39, 45, 46, 48, 52, 58, 59, 62

Behavioral Inhibition System (BIS): 1, 2, 6, 7, 10, 17, 18, 21, 29, 33, 34, 42, 43, 47, 49, 50, 55, 56, 57, 60, 61, 63, 64

Behavioral Approach System (BAS):

- Reward Interest (RI): 11, 13, 14, 15, 26, 32, 35
- Goal-Driven Persistence (GDP): 5, 12, 20, 31, 41, 54, 65
- Reward Reactivity (RR): 3, 4, 8, 16, 23, 24, 25, 30, 36, 37
- Impulsivity (I): 22, 27, 28, 38, 40, 44, 51, 53

Bibliography

- Abdelrahman, Yomna et al. (2017). “Cognitive heat: exploring the usage of thermal imaging to unobtrusively estimate cognitive load”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1.3, p. 33.
- Adan, Ana (1991). “Influence of morningness-eveningness preference in the relationship between body temperature and performance: A diurnal study”. In: *Personality and Individual Differences* 12.11, pp. 1159–1169.
- Adan, Ana et al. (2012). “Circadian typology: a comprehensive review”. In: *Chronobiology international* 29.9, pp. 1153–1175.
- Agrigoroaie, Roxana, Francois Ferland, and Adriana Tapus (2016). “The ENRICHME Project: Lessons Learnt from a First Interaction with the Elderly”. In: *Social Robotics: 8th International Conference, ICSR 2016, Kansas City, MO, USA, November 1-3, 2016 Proceedings*. Springer, pp. 735–745.
- Agrigoroaie, Roxana and Adriana Tapus (2017a). “Contactless Physiological Data Analysis for Users Quality of Life Improving by using a Humanoid Social Robot”. In: *19th Conference on Image Analysis and Processing, ICIAP 2017, Catania, Italy, September 11-15, 2017 Proceedings*. Catania, Italy: Springer.
- (2017b). “Influence of Robot’s Interaction Style on Performance in a Stroop Task”. In: *Accepted in the 9th International Conference on Social Robotics*.
- Agrigoroaie, Roxana, Dan Ciocirlan, and Adriana Tapus (2017c). “News Application Adaptation based on User Sensory Profile”. In: *International Conference on Social Robotics*. Springer, pp. 708–717.
- Agrigoroaie, Roxana and Adriana Tapus (2018a). “‘Oh! I am so sorry!’: Understanding User Physiological Variation while Spoiling a Game Task”. In: *The 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems*.
- (2018b). “Physiological Differences Depending on Task Performed in a 5-Day Interaction Scenario Designed for the Elderly: A Pilot Study”. In: *International Conference on Social Robotics*. Springer, pp. 192–201.
- (2018c). “The outcome of a week of intensive cognitive stimulation in an elderly care setup: a pilot test”. In: *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, pp. 814–819.
- (2019). “Cognitive Performance and Physiological Response Analysis”. In: *International Journal of Social Robotics*. ISSN: 1875-4805. DOI: [10.1007/s12369-019-00532-z](https://doi.org/10.1007/s12369-019-00532-z). URL: <https://doi.org/10.1007/s12369-019-00532-z>.
- Aliakseyeu, Dzmitry et al. (2011). “Exploring interaction strategies in the context of sleep”. In: *IFIP Conference on Human-Computer Interaction*. Springer, pp. 19–36.
- Amos, B et al. (2016). “OpenFace: Face Recognition with Deep Neural Networks”. In: *GITHUB*. URL: <http://github.com/cmusatyalab/openface>.
- Bandura, Albert (1999). “Social cognitive theory of personality”. In: *Handbook of personality: Theory and research*, pp. 154–196.
- Bandura, Albert and Forest J Jourden (1991). “Self-regulatory mechanisms governing the impact of social comparison on complex decision making.” In: *Journal of personality and social psychology* 60.6, p. 941.

- Barrick, Murray R and Michael K Mount (1991). "The big five personality dimensions and job performance: a meta-analysis". In: *Personnel psychology*.
- Bartneck, Christoph and Jodi Forlizzi (2004). "A design-centred framework for social human-robot interaction". In: *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759)*. IEEE, pp. 591–594.
- Bastien, Célyne H, Annie Vallières, and Charles M Morin (2001). "Validation of the Insomnia Severity Index as an outcome measure for insomnia research". In: *Sleep medicine 2.4*, pp. 297–307.
- Bentivoglio, Anna Rita and et.al. (1997). "Analysis of blink rate patterns in normal subjects". In: *Movement disorders*.
- Blatter, Katharina and Christian Cajochen (2007). "Circadian rhythms in cognitive performance: methodological constraints, protocols, theoretical underpinnings". In: *Physiology & behavior 90.2-3*, pp. 196–208.
- Bonnet, Michael H and Donna L Arand (2010). "Hyperarousal and insomnia: state of the science". In: *Sleep medicine reviews 14.1*, pp. 9–15.
- Brown, Catana et al. (2001). "The adult sensory profile: Measuring patterns of sensory processing". In: *American Journal of Occupational Therapy*.
- Brown, Catana and Winnie Dunn (2002). "Adult-Adolescent sensory profile: Self Questionnaire". In: *Pearson Clinical*.
- Cacioppo, John T, Louis G Tassinary, and Gary Berntson (2007). *Handbook of psychophysiology*. Cambridge University Press.
- Campagne, A, T Pebayle, and Alain Muzet (2005). "Oculomotor changes due to road events during prolonged monotonous simulated driving". In: *Biological psychology*.
- Cassell, Justine and Timothy Bickmore (2000). "External manifestations of trustworthiness in the interface". In: *Communications of the ACM 43.12*, pp. 50–56.
- Charles, Rebecca L and Jim Nixon (2019). "Measuring mental workload using physiological measures: A systematic review". In: *Applied Ergonomics 74*, pp. 221–232.
- Cheson, Ashley, Jerry B Richards, and Harriet de Wit (2007). "Effects of sleep deprivation on impulsive behaviors in men and women". In: *Physiology & Behavior*, pp. 79–587.
- Choe, Eun Kyong et al. (2011). "Opportunities for computing technologies to support healthy sleep behaviors". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 3053–3062.
- Christensen, J, Michael Væth, and Ann Wenzel (2012). "Thermographic imaging of facial skin—gender differences and temperature changes over time in healthy subjects". In: *Dentomaxillofacial Radiology 41.8*, pp. 662–667.
- Christos, Kouroupetroglou et al. (2017). "Interacting with Dementia: The MARIO Approach". In: *AAATE Congress*.
- Colquhoun, WP (1960). "Temperament, inspection efficiency, and time of day". In: *Ergonomics 3.4*, pp. 377–378.
- Conners, C Keith et al. (2000). "Conners' continuous performance Test II (CPT II v. 5)". In: *Multi-Health Syst Inc 29*, pp. 175–96.
- Corr, Philip and A. Cooper (2016). "The Reinforcement Sensitivity Theory of Personality Questionnaire: Development and validation." In: *Psychological Assessment*.
- Cruz-Maya, A. and A. Tapus (2018). "Negotiating with a robot: Analysis of Regulatory Focus Behavior". In: *Proceedings of the IEEE International Conference on Robotics and Automation*.
- Cylkowska-Nowak, Mirosława et al. (2015). "The robot in elderly care". In: *Psychology and psychiatry, sociology, health care and education-SGEM 2015*.

- Dalal, Navneet and Bill Triggs (2005). “Histograms of oriented gradients for human detection”. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 1. IEEE, pp. 886–893.
- Davis, Mark H (1983). “Measuring individual differences in empathy: Evidence for a multidimensional approach.” In: *Journal of personality and social psychology*.
- Di Nuovo, Alessandro and et.al. (2017). “The multi-modal interface of Robot-Era multi-robot services tailored for the elderly”. In: *Intelligent Service Robotics*.
- Dinges, David F et al. (1997). “Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 hours per night”. In: *Sleep* 20.4, pp. 267–277.
- Dondrup, C et al. (2015). “Real-time multisensor people tracking for human-robot spatial interaction”. In: *Workshop on Machine Learning for Social Robotics at International Conference on Robotics and Automation - ICRA*.
- Dunn, Winnie (1994). “Performance of typical children on the sensory profile: An item analysis”. In: *American Journal of Occupational Therapy*.
- (1997). “The impact of sensory processing abilities on the daily lives of young children and their families: A conceptual model”. In: *Infants and young children*.
- Durmer, Jeffrey S and David F Dinges (2005). “Neurocognitive consequences of sleep deprivation”. In: *Seminars in neurology*. Vol. 25. 01, pp. 117–129.
- Ehrlich, Stefan et al. (2014). “When to engage in interaction—And how? EEG-based enhancement of robot’s ability to sense social signals in HRI”. In: *Humanoid Robots (Humanoids), 2014 14th IEEE-RAS International Conference on*. IEEE, pp. 1104–1109.
- Elfes, Alberto (1989). “Using occupancy grids for mobile robot perception and navigation”. In: *Computer* 22.6, pp. 46–57.
- Eysenck, H. J. (1991). “Dimensions of Personality: 16, 5 or 3? - Criteria for a Taxonomic Paradigm”. In: *Personality and Individual Differences*.
- Eysenck, Hans Jurgen and Sybil Bianca Giuletta Eysenck (1975). *Manual of the Eysenck Personality Questionnaire (junior and adult)*. Hodder and Stoughton.
- Eysenck, H.J (1967). *The biological basis of personality*. Springfield.
- Eysenck, HJ (1983). “Psychophysiology and personality: Extraversion, neuroticism and psychoticism”. In: *Physiological correlates of human behavior* 3, pp. 13–30.
- Eysenck, H.J. and S.B.G Eysenck (1976). *Psychoticism as a Dimension of Personality*. Hodder and Stoughton, London.
- Faur, Caroline, Jean-Claude Martin, and Céline Clavel (2017). “Measuring chronic regulatory focus with proverbs: The developmental and psychometric properties of a French scale”. In: *Personality and Individual Differences*.
- Ferland, François and Adriana Tapus (2017). “Crowd sourcing ‘approach behavior’ control parameters for human-robot interaction”. In: *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, pp. 785–790.
- Folstein, Marshal F, Susan E Folstein, and Paul R McHugh (1975). ““Mini-mental state”: a practical method for grading the cognitive state of patients for the clinician”. In: *Journal of psychiatric research* 12.3, pp. 189–198.
- Fortier-Brochu, Émilie et al. (2012). “Insomnia and daytime cognitive performance: a meta-analysis”. In: *Sleep medicine reviews* 16.1, pp. 83–94.
- Fröberg, Jan E (1977). “Twenty-four-hour patterns in human performance, subjective and physiological variables and differences between morning and evening active subjects”. In: *Biological psychology* 5.2, pp. 119–134.
- Fukuda, Kyosuke (2001). “Eye blinks: new indices for the detection of deception”. In: *International Journal of Psychophysiology*.

- Goldberg, L.R. (1990). "An Alternative 'Description of Personality': the Big-Five Factor Structure". In: *Journal of Personality and Social Psychology*.
- Gray, J.A. and N. McNaughton (2000). *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system*.
- Hall, Edward Twitchell (1966). *The hidden dimension*. Doubleday & Co.
- Horne, J.A. and O Östberg (1976). "A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms". In: *International Journal of Chronobiology*.
- Horne, JA, CG Brass, and AN Petitt (1980). "Circadian performance differences between morning and evening 'types'". In: *Ergonomics* 23.1, pp. 29–36.
- Ioannou, Stephanos et al. (2014a). "Proximity and gaze influences facial temperature: a thermal infrared imaging study". In: *Frontiers in psychology*.
- Ioannou, Stephanos, Vittorio Gallese, and Arcangelo Merla (2014b). "Thermal Infrared Imaging in Psychophysiology: Potentialities and limits". In: *Psychophysiology*.
- Kahneman, Daniel (1973). *Attention and effort*. New Jersey: Prentice-Hall.
- Kazemi, Vahid and Josephine Sullivan (2014). "One millisecond face alignment with an ensemble of regression trees". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- King, D. (2009). "Dlib-ml: A Machine Learning Toolkit". In: *Journal of Machine Learning Research*. Vol. 10, pp. 1755–1758.
- Kostavelis, Ioannis et al. (2015). "Ramcip: Towards a robotic assistant to support elderly with mild cognitive impairments at home". In: *International Symposium on Pervasive Computing Paradigms for Mental Health*. Springer, pp. 186–195.
- Kramer, Arthur F (1991). "Physiological metrics of mental workload: A review of recent progress". In: *Multiple-task performance*, pp. 279–328.
- Lisetti, Christine Lætitia and Fatma Nasoz (2004). "Using noninvasive wearable computers to recognize human emotions from physiological signals". In: *EURASIP Journal on Advances in Signal Processing*.
- Liu, Changchun and et.al. (2008). "Online affect detection and robot behavior adaptation for intervention of children with autism". In: *IEEE transactions on robotics*.
- MacLeod, Colin (1991). "Half a Century of Research on the Stroop Effect: An Integrative Review". In: *Psychological Bulletin*.
- Mahoney, Florence I and Dorothea W Barthel (1965). "Functional evaluation: the Barthel Index: a simple index of independence useful in scoring improvement in the rehabilitation of the chronically ill." In: *Maryland state medical journal*.
- Martins, Gonçalo S, Luis Santos, and Jorge Dias (2015). "The GrowMeUp project and the applicability of action recognition techniques". In: *Third workshop on recognition and action for scene understanding (REACTS)*. Ruiz de Aloza.
- Merten, Matthias et al. (2012). "A mobile robot platform for socially assistive home-care applications". In: *ROBOTIK 2012; 7th German Conference on Robotics*. VDE, pp. 1–6.
- Metta, Giorgio et al. (2008). "The iCub humanoid robot: an open platform for research in embodied cognition". In: *Proceedings of the 8th workshop on performance metrics for intelligent systems*. ACM, pp. 50–56.
- Mitler, Merrill M (2004). "Maintenance of Wakefulness Test". In: *Sleep Deprivation*. CRC Press, pp. 53–66.
- Mitsunaga, Noriaki and et.al. (2006). "Robot behavior adaptation for HRI based on policy gradient reinforcement learning". In: *Journal of the Robotics Society of Japan*.

- Mower, Emily et al. (2007). “Investigating implicit cues for user state estimation in human-robot interaction using physiological measurements”. In: *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*. IEEE, pp. 1125–1130.
- Nakatsu, Ryohei (1998). “Toward the creation of a new medium for the multimedia era”. In: *Proceedings of the IEEE* 86.5, pp. 825–836.
- Nani, Maria et al. (2010). “MOBISERV: an integrated intelligent home environment for the provision of health, nutrition and mobility services to the elderly”. In: Natale, Vincenzo, Antonella Alzani, and PierCarla Cicogna (2003). “Cognitive efficiency and circadian typologies: a diurnal study”. In: *Personality and Individual differences* 35.5, pp. 1089–1105.
- Nebylitsyn, Vladimir Dmitrievich and Jeffrey Alan Gray (1972). *Biological bases of individual behavior*. Academic Press.
- Nourbakhsh, Nargess and et.al. (2012). “Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks”. In: *Proceedings of the 24th Australian Computer-Human Interaction Conference*.
- Nourbakhsh, Nargess, Yang Wang, and Fang Chen (2013). “GSR and blink features for cognitive load classification”. In: *IFIP Conference on Human-Computer Interaction*, pp. 159–166.
- Nourbakhsh, Nargess et al. (2017). “Detecting Users’ Cognitive Load by Galvanic Skin Response with Affective Interference”. In: *ACM Transactions on Interactive Intelligent Systems (TiiS)* 7.3, 12:1–12:20.
- Ohayon, Maurice M and Charles F Reynolds III (2009). “Epidemiological and clinical relevance of insomnia diagnosis algorithms according to the DSM-IV and the International Classification of Sleep Disorders (ICSD)”. In: *Sleep medicine* 10.9, pp. 952–960.
- Or, Calvin KL and Vincent G Duffy (2007). “Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement”. In: *Occupational Ergonomics* 7.2, pp. 83–94.
- Paas, Fred GWC and Jeroen JG Van Merriënboer (1993). “The efficiency of instructional conditions: An approach to combine mental effort and performance measures”. In: *Human factors* 35.4, pp. 737–743.
- Pages, Jordi, Luca Marchionni, and Francesco Ferro (2016). “TIAGo: the modular robot that adapts to different research needs”. In: *International Workshop on Robot Modularity, IROS*.
- Panda, Dr. Keshab. *Robotics and AI improve factories of the future*. <https://www.controleng.com/articles/robotics-and-ai-improve-factories-of-the-future/>. Accessed: 13/05/2019.
- Papoutsakis, K and et.al. (2013). “Developing visual competencies for socially assistive robots: the HOBbit approach”. In: *Proceedings of the 6th International Conference on Pervasive Technologies Related to Assistive Environments*. ACM, p. 56.
- Paul, Michel A et al. (2003). “Impact of melatonin, zaleplon, zopiclone, and temazepam on psychomotor performance”. In: *Aviation, space, and environmental medicine* 74.12, pp. 1263–1270.
- Petrides, Kostantinos V (2009). “Psychometric properties of the trait emotional intelligence questionnaire”. In: *Assessing emotional intelligence*.
- Picard, Rosalind W. and et.al. (2001). “Toward machine emotional intelligence: Analysis of affective physiological state”. In: *IEEE transactions on pattern analysis and machine intelligence*.
- Quigley, M and et.al. (2009). “ROS: an open-source Robot Operating System”. In: *ICRA Workshop on Open Source Software*. Vol. 3, p. 5.

- Rahman, Hamidur et al. (2016). “Real time heart rate monitoring from facial RGB color video using webcam”. In: *SAIS*.
- Randler, C. (2008). “Differences between smokers and nonsmokers in morningness-eveningness”. In: *Social Behavior and Personality*.
- Reeves, Byron and Clifford Ivar Nass (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge university press.
- Renaud, Patrice and Jean-Pierre Blondin (1997). “The stress of Stroop performance: physiological and emotional responses to color-word interference, task pacing, and pacing speed”. In: *International Journal of Psychophysiology*.
- Rodgers, Shannon et al. (2016). “Managing stress, sleep and technologies: an exploratory study of Australian university students”. In: *Proceedings of the 28th Australian Conference on Computer-Human Interaction*. ACM, pp. 526–530.
- Sagonas, Christos et al. (2016). “300 faces in-the-wild challenge: Database and results”. In: *Image and vision computing* 47, pp. 3–18.
- Savitzky, Abraham and Marcel JE Golay (1964). “Smoothing and differentiation of data by simplified least squares procedures.” In: *Analytical chemistry* 36.8, pp. 1627–1639.
- Setz, Cornelia and et.al. (2010). “Discriminating stress from cognitive load using a wearable EDA device”. In: *IEEE Transactions on information technology in biomedicine*.
- Shi, Yu et al. (2007). “Galvanic skin response (GSR) as an index of cognitive load”. In: *CHI’07 extended abstracts on Human factors in computing systems*. New York, NY, USA: ACM, pp. 2651–2656.
- Song, John and Con Stough (2000). “The relationship between morningness–eveningness, time-of-day, speed of information processing, and intelligence”. In: *Personality and Individual differences* 29.6, pp. 1179–1190.
- Sorostinean, Mihaela, François Ferland, and Adriana Tapus (2015). “Reliable Stress Measurement using Face Temperature Variation with a Thermal Camera in Human-Robot Interaction”. In: *15th IEEE-RAS International Conference on Humanoid Robots*.
- Spreng, R Nathan and et.al. (2009). “The Toronto Empathy Questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures”. In: *Journal of personality assessment*.
- Sprent, Peter and Nigel C Smeeton (2000). *Applied nonparametric statistical methods*. Chapman and Hall CRC.
- Stelmack, Robert M. (1981). “The Psychophysiology of Extraversion and Neuroticism”. In: *A Model for Personality*. Ed. by Hans J. Eysenck. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 38–64.
- Stern, John A, Larry C Walrath, and Robert Goldstein (1984). “The endogenous eyeblink”. In: *Psychophysiology*.
- Stroop, J.R. (1935). “Studies of interference in serial verbal reactions”. In: *Journal of Exploratory Psychology*.
- Svetlak, Miroslav and et.al. (2010). “Electrodermal complexity during the Stroop colour word test”. In: *Autonomic Neuroscience*.
- Tanaka, Yuu and Kiyoshi Yamaoka (1993). “Blink Activity and Task Difficulty”. In: *Perceptual and Motor Skills*.
- Tapus, Adriana, Cristian Tapus, and Maja J Mataric (2007a). “Hands-off therapist robot behavior adaptation to user personality for post-stroke rehabilitation therapy”. In: *Robotics and Automation, 2007 IEEE International Conference on*.

- Tapus, Adriana and et.al. (2007b). "Socially assistive robotics [grand challenges of robotics]". In: *IEEE Robotics & Automation Magazine*.
- Tulen, JHM and et.al. (1989). "Characterization of stress reactions to the Stroop Color Word Test". In: *Pharmacology Biochemistry and Behavior*.
- Turner, J Rick and Douglas Carroll (1985). "Heart Rate and Oxygen Consumption during Mental Arithmetic, a Video Game, and Graded Exercise". In: *Psychophysiology*.
- Wada, Kazuyoshi et al. (2003). "Effects of robot assisted activity to elderly people who stay at a health service facility for the aged". In: *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)(Cat. No. 03CH37453)*. Vol. 3. IEEE, pp. 2847–2852.
- Walters, M and et.al. (2005). "The influence of subjects' personality traits on personal spatial zones in a human-robot interaction experiment". In: *IEEE International Workshop on Robot and Human Interactive Communication*.
- Wang, Shangfei et al. (2010). "A natural visible and infrared facial expression database for expression recognition and emotion inference". In: *IEEE Transactions on Multimedia*.
- Watson, David and L Clark (1988). "The Positive and Negative Affect Schedule: PANAS Questionnaire". In: *Journal of personality and social psychology*.
- Woods, Sarah and et.al. (2007). "Are robots like people?: Relationships between participant and robot personality traits in human-robot interaction studies". In: *Interaction Studies*.
- Yanco, Holly A, Jill L Drury, and Jean Scholtz (2004). "Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition". In: *Human Computer Interaction* 19.1-2, pp. 117–149.
- Yu, B et al. (1999). "Estimate of the Optimum Cutoff Frequency for the Butterworth Low-Pass Digital Filter". In: *Journal of Applied Biomechanics*. Vol. 15.

Titre : Explorer la relation entre l'échelle de typologie circadienne, la performance cognitive et l'état physiologique dans différents scénarios d'interaction homme-robot.

Mots clés : robotique d'assistance, interaction homme-machine, état physiologique

Résumé : Les systèmes de robotique sociale sont de plus en plus présents dans nos vies. Ce ne sont plus des entités isolées, mais on s'attend à ce qu'ils soient capables d'interagir et de communiquer avec les humains. Ils doivent respecter les normes comportementales attendues par les humains avec qui les systèmes robotiques sont en interaction. L'une des principales pistes de recherche dans le domaine de la robotique sociale est représentée par la conception d'une interaction naturelle entre un robot social et un individu. Plus spécifiquement, cette interaction devrait prendre en considération le profil de l'individu, l'état émotionnel, l'état physiologique et l'humeur, entre autres.

Dans cette thèse, nous explorons la relation qui existe entre l'échelle de typologie circadienne, la performance cognitive et l'état physiologique au cours de différents scénarios d'interaction homme-robot. L'administration de différents questionnaires psychologiques permet de déterminer le profil d'un individu. En outre, à l'aide de différents capteurs (par exemple,

GSR, caméra thermique), de multiples méthodologies ont été développées pour déterminer l'état physiologique d'un individu. Plus spécifiquement, la variation de la température faciale, le clignotement des yeux et la réponse galvanique de la peau ont été étudiés.

Plusieurs scénarios d'interaction homme-robot ont été conçus afin de tester le système développé. L'impact de l'empathie a également été étudié. En outre, le système développé a été testé avec succès dans deux environnements réels, avec deux populations vulnérables. La première application d'assistance est représentée par le projet de recherche EU H2020 ENRICHME, dans lequel un robot a été développé pour les personnes âgées atteintes d'un trouble cognitif léger. La deuxième population vulnérable est constituée d'individus souffrant de différents troubles du sommeil. Nous pensons que cette thèse représente une étape importante dans la compréhension de l'état physiologique de l'individu et est liée à la performance cognitive.

Title : Exploring the relationship between morningness-eveningness, cognitive performance and the internal physiological state in different human-robot interaction scenarios.

Keywords : assistive robotics, human-robot interaction; physiological internal state

Abstract : Social robotic systems are more and more present in our everyday lives. They are no longer isolated entities, but instead, they are expected to be capable of interacting and communicating with humans. They have to follow the behavioral norms that are expected by the individuals the robotic systems are interacting with. One of the main research directions in the field of social robotics is represented by the design of a natural interaction between a social robot and an individual. More specifically, this interaction should take into consideration the profile of the individual, the emotional state, the physiological internal state, and the mood, among others.

In this thesis it is explored the relationship that exists between morningness-eveningness, cognitive performance, and the internal physiological state during different human-robot interaction scenarios. By administering different psychological questionnaires, the profile of an individual can be determined. Moreover, with the help of different sensors (e.g., GSR, thermal ca-

mera), multiple methodologies were developed to determine the internal physiological state of an individual. More specifically, the facial temperature variation, the blinking, and the galvanic skin response were investigated.

Several human-robot interaction scenarios have been designed in order to test the developed system. The impact of empathy was also investigated. Furthermore, the developed system was successfully tested in two real-world environments, with two vulnerable populations. The first assistive application is represented by the ENRICHME EU H2020 research project, where a personal robot was developed for the elderly with mild cognitive impairment. The second vulnerable population consists of individuals suffering from different sleep disorders. We believe that this thesis represents an important step in understanding how the physiological internal state of an individual is related to cognitive performance, and to the user profile of that individual.

