A Multimodal Approach for Identification, Analysis, and Comparison of Human Emotional and Behavioral Patterns between Human - Human and Human - Technology Interaction

Baka, Evangelia

UNIVERSITÉ DE GENÈVE

	FACULTÉ DES SCIENCES
Département d'informatique	Professeur José Rolim
	FACULTÉ D'ÉCONOMIE ET DE MANAGEMENT
Département de systeme d'information	Professeur Nadia Magnenat Thalmann

A Multimodal Approach for Identification, Analysis, and Comparison of Human Emotional and Behavioral Patterns between Human – Human and Human – Technology Interaction

THÈSE

présentée à la Faculté des sciences de l'Université de Genève pour obtenir le grade de Docteur ès sciences, mention interdisciplinaire

par

Evangelia BAKA

de

Athènes (Grèce)

Thèse Nº

GENÈVE

Atelier d'impression ..

2022

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor Prof. Nadia Magnenat Thalmann for her constant support, guidance, and supervision throughout this Ph.D. work. Being a member of MIRALab allowed me to enhance my knowledge, explore new directions and spread my research interests as well as improve in depth my skills. I also owe a big thank to the jury members Prof. José Rolim (University of Geneva), Prof. George Papagiannakis (University of Crete and FORTH) and Prof. Yiyu Cai (Nanyang Technological University) for their time and effort in reviewing my manuscript.

I am also grateful to my colleagues at MIRALab, Nedjma Cadi, Marlène Arevalo, Simon Senecal and Evropi Stefanidi for being there, supporting me and making this experience pleasant and significant. A big thank also to Dr. Guillaume Chanel, for facilitating my first research experiment at the University of Geneva by providing us the with EEG device. I need also to thank all the people of IMI at the Nanyang Technological University in Singapore, and especially Nidhi Mishra, for their continuous support, collaboration, and share of knowledge without which I would have never completed this work.

The presented work was supported by several projects: the European Union's Horizon 2020 NOTRE, VIMM, and MINGEI. I would like to thank all the people I met during these exchanges for their collaboration but also their hospitality. These experiences helped me broaden my research topic and define my research goals.

Finally, I would like to thank all my close friends for being there and my parents for their unconditional support, patience, and for believing in me. Last but not least, without the love and the support of my husband this journey would have never been possible. A final special thank to my little son who came during this journey and filled me with motivation.

ABSTRACT

Our everyday life includes more than one social interaction, which are based mainly on human communication. The latter consists of human emotional and behavioral informations, expressed by voice, facial expressions or body movements. However, a social interaction is not limited between humans and thus, Human-Computer and Human-Robot Interactions (HRI) are meeting their golden point of research. The multidisciplinary area of computer graphics, neuroscience, psychology, and artificial intelligence is trying to explore, decipher, decode and model the human emotional behaviors and expressions to enhance the field of social interactions between humans and technology. Toward this direction, the goal of this research is to delve deeper into the features and the nature of Human – Human (H-H) and Human – NonHuman (H-NH) social interactions, providing a detailed validated assessment of how humans react towards technology but also how the latter affects humans.

In the first place, we examined the effect of the well-used virtual environments in the brain, identifying possible differences in perception between the virtual and the real world. An EEG device was used to capture participants' brain activity in different brain areas examining motor, cognitive and other function of the users and a questionnaire was used to evaluate psychological factors and the sense of presence. Our results enhanced the current literature by revealing new brain states involved in virtual environments, like frontal theta state and centra alpha state and they highlighted the importance of the graphics content revealing a difference in the occipital area.

The second part of our work is concentrated on the broad area of HRI, consisting of two different experiments. The first one concerned the identification of the effects human-humanoid interaction can have on human emotional states and behaviors, through a physical interaction with a robot, an identical human and a human. This research was supported by EEG and audio recordings as well as a questionnaire indicating that the human brain does understand visually and auditorily the difference between a robot and an identical human but the levels of concentration and motivation remain higher during HRI.

The second experiment led us to multidisciplinary in-depth documentation, analysis, and comparison between H-H and H-NH interactions recording brain activity, muscles activity, body movements, voice, and emotional states. A dataset was created with 40 participants interacting with three different agents (human, virtual human and a robot) under the same scenario. The robot was also tested under four different roles. Human emotional and behavioral patterns were extracted and compared, providing valuable insights

regarding the role of robots and the effort for humanization. The role of physical presence was also assessed. Up-to-date researches show us that the need is focused on designing social agents in a more human-like way behaviorwise and not in terms of appearance and our study complemented the above concluding that it is the reactions of the agents that trigger the different human responses and not the appearance alone. Lastly, we developed a model that can recognize and classify the human voice based on the nature of the interlocutor with a score of 82%.

This thesis bridge the gap among studies that have examined the role of human – likeness in nonhuman agents, studies that have examined human-human interactions alone and the ones that have analyzed and compared human reactions during nonhuman social interactions. Further research with more extended studies is required to shed more light to this broad area of H-NH interaction, revealing human reactions that can guide the design of nonhuman agents, can elicit better cognitive and emotional responses and ensure a higher level of engagement, respecting the human needs.

RÉSUMÉ

Notre vie quotidienne est rythmée par plus d'une interaction dont la base repose principalement sur la communication humaine. Cette dernière est constituée d'informations émotionnelles et comportementales exprimées par la voix, les expressions faciales ou les mouvements corporels. Cependant, une interaction sociale n'est pas limitée aux humains et, par conséquent, les interactions Humain-Ordinateur et Humain-Robot connaissent une croissance exponentielle du nombre de recherches s'y attardant. Les domaines de l'infographie, de la neuroscience, de la psychologie et de l'intelligence artificielle tentent d'explorer, de déchiffrer et de modéliser les comportements et expressions émotionnels des personnes afin de renforcer les interactions sociales que nous avons avec la technologie. Dans le prolongement de ceci, l'objectif de cette recherche est de plonger plus profondément dans les caractéristiques et la nature des échanges sociaux entre humains et lorsque la personne interagit avec les non humains. Cette thèse fournit une évaluation détaillée et validée de la réaction des êtres humains face à la technologie, mais également de la façon dont celle-ci affecte les premiers.

Premièrement, nous avons examiné les effets induits par les environnements virtuels dans le cerveau. Nous avons donc identifié les différences possibles de perception entre les mondes virtuels et réels. Un dispositif EEG a été utilisé pour capturer l'activité cérébrale des participants sur plusieurs zones en examinant les fonctions motrices, cognitives et d'autres types. Nous avons également créé un questionnaire évaluant les facteurs psychologiques et le sentiment de présence. Nos résultats améliorent la littérature actuelle en révélant de nouveaux signaux cérébraux lorsque ce dernier est confronté à un environnement virtuels. Parmis ces signaux, nous pouvons citer l'état thêta frontal et alpha central. Finalement, nos conclusions soulignent l'importance des contenus graphiques révélant une différence dans la zone occipitale.

Dans un second temps, notre travail s'est concentré autour de deux expériences basées sur le vaste domaine de l'IRH. La première a identifié les effets que l'interaction humain-humanoïde peut avoir sur les émotions et les comportements humains. Trois types de contreparties ont été impliquées : un robot, son clone humain et un humain. Cette recherche s'appuie sur des enregistrements EEG et audio ainsi qu'un questionnaire indiquant que le cerveau de l'humain comprend bien visuellement et auditivement la différence entre un robot et un humain identique. Cette expérience met également en exergue des niveaux de concentration et de motivation plus élevés pendant l'IRH.

La deuxième expérience nous a conduit à une documentation, une analyse et une comparaison multidisciplinaires approfondies des interactions H-H et H-NH en enregistrant l'activité cérébrale et musculaire, les mouvements du corps, la voix et les états émotionnels. Un ensemble de données a été créé sur la base de 40 participants interagissant avec trois agents différents (humain, avatar et robot) dans des scénarios identiques. Le robot a également été testé sous quatre rôles différents. Les schémas émotionnels et comportementaux humains ont été extraits et comparés fournissant de précieuses indications sur le rôle des robots et l'effort d'humanisation. Le rôle de la présence physique a également été évaluée. De récentes recherches montrent que l'effort d'humanisation des agents sociaux doit se concentrer sur l'intégration de comportements plus humains et non sur l'apparence de ceux-ci. Notre étude complète ce postulat en affirmant que ce sont les réactions des agents qui déclenchent certaines réponses humaines et non l'apparence seule. Enfin, nous avons développé un modèle capable de reconnaître et classifier la voix humaine en fonction de la nature de l'interlocuteur avec une précision de 82%.

Cette thèse comble l'écart entre les études ayant analysé le rôle de l'humain, les études examinant uniquement les interactions entre humains et celles qui ont comparé et analysé les réactions humaines lors d'interactions sociales non humaines. Plus de recherches avancées sont nécessaires afin de mettre en lumière ce vaste domaine des interactions entre l'humain et le non humain. Afin de susciter de meilleures réactions cognitives et émotionnelles chez l'humain lorsqu'il interagit avec des agents non humains, il est primordial d'intégrer des réactions humaines dans ces mêmes agents non humains. Finalement, cela assurera un niveau d'engagement plus élevé tout en respectant les besoins humains.

LIST OF PUBLICATIONS

List of publications directly related with this thesis:

Peer-reviewed Conferences:

- Baka E., Mishra N., Magnenat-Thalmann N. (2022) "Social robots and Digital humans as Job Interviewers: A study of humans' reactions towards a more natural interaction", International Conference on Human Computer Interaction (HCII) 2022 (Submitted and Accepted)
- Baka E., Vishwanath A., Mishra N., Vleioras G, Magnenat Thalmann N.(2019), "Am I talking to a Human or a Robot?": A preliminary study of Human's perception during Human – Humanoid interaction and its effects in cognitive and emotional states. M. Gavrilova et al. (Eds.): CGI 2019, LNCS 11542, pp. 240–252, 2019. https://doi.org/10.1007/978-3-030-22514-8_20
- Baka E., Stavroulia KE., Magnenat-Thalmann N., Lanitis A (2018), An EEG-based evaluation for Comparing the sense of presence between Virtual and Physical Environments. In proceedings of Computer Graphics International 2018 (CGI 2018), ACM, New York, USA, 10 pages. https://doi.org/10.1145/3208159.3208179
- Stavroulia KE, Baka E., Lanitis A., Magnenat – Thalmann N. (2018), Designing a virtual environment for teacher training: Enhancing presence and empathy. In proceedings of Computer Graphics International 2018 (CGI 2018), ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3208159.3208177

Book chapters:

- Baka, E., & Thalmann, N. M. (2021). Human—Technology Interaction: The State-of-the-Art and the Lack of Naturalism. In *Intelligent Scene Modeling and Human-Computer Interaction* (pp. 221-239). Springer, Cham.

- Mishra, N., Baka, E., & Thalmann, N. M. (2021). Exploring Potential and Acceptance of Socially Intelligent Robot. In *Intelligent Scene Modeling and Human-Computer Interaction* (pp. 259-282). Springer, Cham. https://doi.org/10.1007/978-3-030-71002-6_15

Peer-reviewed Journals:

- Baka E., Mishra N., Magnenat Thalmann N., Frantzidis C., Vleioras G. (2022) Human, Avatar or a Robot? A multimodal analylsis towards uncovering human's perception and reactions to social interactions (Submitted to Frontiers in Psychology)

List of publications indirectly related with this thesis:

Peer-reviewed Conferences:

- Christofi, M., Baka, E., Stavroulia, K.E., Michael-Grigoriou, D., Lanitis, A. & Magnenat-Thalmann, N. (2018). Studying Levels of Presence in a Virtual Environment Simulating Drug Use in Schools: Effect on Different Character Perspectives. ICAT-EGVE 2018 – 28th International Conference on Artificial Reality and Telexistence (ICAT 2018) and the 23rd Eurographics Symposium on Virtual Environments (EGVE 2018). Limassol, Cyprus.
- Stavroulia, K.E., Baka, E., Christofi, M., Michael-Grigoriou, D., Magnenat-Thalmann, N. & Lanitis, A. (2018). A virtual reality environment simulating drug use in schools: effect on emotions and mood states. In proceedings of International Conference on Information, Communication Technologies in Education, ICICTE 2018, Chania, Greece, 5 and 7 July, 2018.
- Stavroulia, K.E., Baka, E., Lanitis, A., & Magnenat Thalmann, N. (2017). Virtual reality-based learning environments in teacher training: new opportunities and challenges. In the proceedings of the 10th International Conference of Education, Research and Innovation, ICERI2017 Conference (pp. 3562-3571), 16th-18th November 2017, Seville, Spain. ISBN: 978-84-697-6957-7

Peer-reviewed Journals:

- Stavroulia, K. E., Christofi, M., Baka, E., Michael-Grigoriou, D., Magnenat-Thalmann, N., & Lanitis, A. (2019). Assessing the emotional impact of virtual reality-based teacher training. *The International Journal of Information and Learning Technology*. https:// doi.org/10.1108/IJILT-11-2018-0127

CONTENTS

ACKNOWLEDGEMENTS i

ABSTRACT ii

RÉSUMÉ iv

LIST OF PUBLICATIONS vi

LIST OF FIGURES xiv

LIST OF TABLES xviii

Introduction 2

1.1. Human-NonHuman Interaction 2

1.1.1 Research context and motivations 3

Affective signals 3

Nature and design 4

1.2 Objectives and Contributions 7

1.2.1 Limitations 9

1.2.2. Research questions 10

1.3 Potential Applications 11

1.4 Manuscript Organization 12

Related Work 14

2.1 Introduction 14

2.1.1 The role of human likeness 15

2.1.2 The role of embodiment and presence *17*

2.1.3 Other features that can influence human's perception during HCI *19*

Eye gaze and facial expressions 19

Voice and gender 20

2.2 Human-Robot Interaction (HRI) **21**

2.2.1 Social robots and their features *22*

2.2.2 Robots and their roles *25*

Robots as Interviewers 25

Robots as Teachers 26

Robots as Costumer guides 26

Robots as Companions 27

2.3 Humans and Virtual Humans **27**

2.3.1 Conceptualization and Perception of Avatars *28*

2.4 Nonverbal Communication **30**

2.4.1 Part 1 – Encoding process *30*

2.4.2 Part 2 – Decoding process *33*

2.4.3 Mimicry or adjustment? *33*

2.5 Affective Computing and Social Signal Processing **34**

2.5.1 Affect Recognition *35*

2.5.2 Affect Computation *36*

Audio modality 36

Body movements and gestures 37

Physiological and Neurophysiological signals 38

2.6 Modeling, Analysis, and Synthesis of Human Behavior **40**

2.7 Summary and Discussion **49**

Interaction in Virtual and Physical Environments 52

3.1 Introduction **52**

3.2 Experimental Design **52**

3.2.1 Participants *54*

3.3 Data collection and Analysis **54**

3.3.1 EEG recordings and Analysis *54*

Brain signaling and Brain waves 55

3.3.2 Psychometric data *57*

3.3.3 Statistics *57*

3.4 Results: Evaluation of Human Perception between Virtual and Physical Environments **57**

3.4.1 Effects of VR in specific regions of the brain, influencing behavioral, motor or other functions *58*

3.4.2 The role of the VR design *60*

3.4.3 Brain adaptation time to a VE *61*

3.5 Summary and Discussion **62**

Human – Robot Interaction **66**

4.1 Introduction **66**

4.2 Experimental Design **66**

4.2.1 Participants *68*

4.3 Data Collection and Analysis **68**

4.3.1 EEG recordings and analysis *68*

4.3.2 Dialog and audio analysis *69*

The social robot Nadine's architecture 70

4.3.3 Psychometric data *70*

4.3.4 Statistical analysis *71*

4.4 Human Perception during Human-humanoid Interaction and its Effects in Human Cognitive and Emotional States **71**

4.4.1 Brain activity during human-humanoid interaction *71*

4.4.2 Audio data and human perception between HH and HR Interactions *74*

4.4.3 Differences in emotions and motivation when interacting with a human and an identical robot.....75

4.5 Summary and Discussion 77

Human – Nonhuman Interaction 81

5.1 ..... Introduction 81

5.2 ..... Human Behaviors and Reactions during H-NH Interactions 82

5.2.1 Experimental design 82

5.2.2 Participants 85

5.2.3 Data acquisition and analysis 86

EEG recordings and analysis 86

Motion Captures and analysis 88

EMG recordings and analysis 90

Audio recordings and analysis 91

Psychometric data 95

Statistical analysis 96

5.2.4 Human reactions and behaviors during H-H and H-NH Interactions.....96

Voice and body reactions 96

Reactions' correlations.....117

Role of gender and ethnicity 119

5.2.5 Discussion 121

Human reactions 121

Role of gender and ethnicity 126

Conclusion 126

5.3 ..... Potential and Acceptance of Social Robots 127

5.3.1 Experimental design 127

5.3.2 Data acquisition and analysis 128

EEG recordings and analysis 128

Motion data 128

Psychometric data 129

5.3.3 Human perception in HRI under different roles *129*
5.3.3.1 Participant's reactions and preference over four predefined robot roles 129
5.3.3.2 Overall attitude towards robots 135
5.3.4 Discussion *136*
5.4 General Discussion **139**
Conclusion **143**
6.1 Discussion **143**
6.2 Contributions **145**
6.3 Limitations and Future Research **146**
APPENDIX A **148**
A1. NOTRE: Preliminary work on EEG and HR recordings and analysis *148*
A1.1 Conclusion and Contribution 149
A2. MINGEI: Face and Body reconstruction *150*
A2.1 Conclusion and Contribution *151*
BIBLIOGRAPHY **152**

LIST OF FIGURES

Figure 1.1 The humanoid robotic head Eva 5

Figure 2.1 The flow of the literature review 14

Figure 2.2 The uncanney valley 16

Figure 2.3 A framework of nonverbal encoding cues created by Hall et al, 2019..... 32

Figure 2.4 The most recent table, created by Giger et al. [207], expressing the thoughts and concerns of humanizing social robots 43

Figure 2.5 The conceptualization of robotic psychology as modeled by Stock et al. 44

Figure 2.6 Diagram depicted the emotional contagion in HRI, as designed by Stock et al. 45

Figure 3.1 Screenshots of the environments used for the experiment. 53

Figure 3.2 LEFT: A participant weating the headcap with some EEG electrodes on. RIGHT: Regions of Interest, separated according to the brain areas we wanted to examine..... 55

Figure 3.3 ABOVE: The frequencies and the characteristics of the five basic brain waves. BELOW: Samples of the five basic brain waves in time domain 56

Figure 3.4 Power spectra observed in each of the 10 ROIs, in response to each environment..... 59

Figure 3.5 The duration of time needed for the adaptation of the brain to the new state 61

Figure 3.6 Summary of the results extracted from the EEG data and the questionnaire.**Error! Bookmark not defined.**

Figure 4.1 Participants during the three types of interaction. 67

Figure 4.2 Regions of Interest (ROIs) used for this study. 69

Figure 4.3 Nadine's architecture 70

Figure 4.4 Power spectra observed in each of the 5 ROIs, in response to each case. 73

Figure 4.5 Mean of frequencies for each brain area for both hemispheres in three conditions 73

Figure 4.6 Differences in frequencies of each case for the two hemispheres 74

Figure 4.7 Mean values for Pitch, Intensity and duration of the interaction for each condition, derived from the audio analysis 75

Figure 4.8 Participants' emotional states for each condition. 76

Figure 4.9 Summary of the results of the brain activity for the examined five brain areas. 78

Figure 5.1 LEFT: case VH where the participant interacted with Nicole, the virtual human. RIGHT: the nonhuman agents used in our experiment 82

Figure 5.2 The setup of our work 83

Figure 5.3 The flow of our research protocol 85

Figure 5.4 Example of the visual manual inspection conducted in EEGLab 86

Figure 5.5 The 25 points examined with Kinect 88

Figure 5.6 LEFT: The four muscles' positions where we put the EMG electrodes. RIGHT: An example of a participant wearing the physiological equipment (EEG and EMG) while interacting with Nadine. 91

Figure 5.7 LEFT: Audio sample of a female participant interacting with the social robot Nadine in Praat software. RIGHT: Sample of the Nadine's voice. Up: the voice signal with the voice breaks. Down: the spectrogram. 93

Figure 5.8 Relative energies for the five brain states in the six brain areas for the three interactions 99

Figure 5.9 Differences in brain states per interaction. 100

Figure 5.10 Changes in movements' range during the three interactions. 101

Figure 5.11 Percentages of emotions extracted from the body movement captured by Kinect in the three interactions. 103

Figure 5.12 Example of a participant's spectogram extracted from the wavelet analysis for the four muscles in each side and for each interaction. 106

Figure 5.13 Comparison of the features between humans' and agents' responses. 112

Figure 5.14 Emotions extracted from the voice signal for the three interactions. 113

Figure 5.15 Confusion matrix 114

Figure 5.16 Results from the KNN classification 114

Figure 5.17 The dominant emotions in H-NA and H-H respectively that presented statistically significant differences with the other conditions. 116

Figure 5.18 The scores of participants' perception towards all the agents. 117

Figure 5.19 The significant differences found in features in function of the gender and/or ethnicity..... 119

Figure 5.20 Mixed ANOVA with ethnicity as between-subject variable for Biceps and Trapezius for the three interactions. 120

Figure 5.21 Response time between the participant and the agent. 123

Figure 5.22 Summary of the human reactions per interaction for each modality 125

Figure 5.23 Example of a participant interacting with Nadine in the role of the customer guide 128

Figure 5.24 The five Regions of Interest (ROIs) used for both parts of the experiment. 128

Figure 5.25 The three channels recorded in occipital area showing the dominant brain state for each role in the power spectrum. 130

Figure 5.26 The average of frequency in the five brain areas for the four roles of Nadine. 131

Figure 5.27 Discreet emotions extracted from body skeleton movements through Kinect V2. 132

Figure 5.28 Negative and positive emotional states as described by our participants for the four roles of Nadine. 134

Figure 5.29 Results from our questionnaire regarding the preference of the four roles. 135

Figure 8.1 The VE as seen from the different perspectives. 149

Figure 8.2 Dominant brain frequency for healthy student (Left) and for teacher (Right) perspectives. The diagram was constructed after ICA was applied. 149

Figure 8.3 Image of the scanner in face position 150

Figure 8.4 Face and Body reconstruction 151

Figure 8.5 Face wrapping 151

LIST OF TABLES

Table 2.1 Examples of social robots from several domains in a chronological order 24

Table 2.2 The most commonly used multimodal datasets up to now on realistic human social behavior. 40

Table 2.3 Studies on human perception and behavior under several types of interaction during the last decade..... 46

Table 4.1 Thematic areas used for facilitation of the discussion during the three interaction..... 67

Table 5.1 The thematic areas used for the job interview scenario and the questions for each one of them as posed for case of Nadine robot. 84

Table 5.2 Demographic data of the study sample 85

Table 5.3 The five body parts we examined and the points used for each one..... 89

Table 5.4 Acoustic/Prosodic and Conversational Features used..... 92

Table 5.5. Scores of dimensionality reduction methods 94

Table 5.6 Comparison of the different models..... 95

Table 5.7 Differences in the brain states of each brain area among interactions..... 97

Table 5.8 Brain states per interaction..... 99

Table 5.9 Differences in the range of movement between the conditions H-NA, H-VH and HH..... 102

Table 5.10 Mean Values (SD) for the mean frequency (Fmean) and the root mean square (RMS) for both sides of the four muscles per interaction..... 104

Table 5.11 Summary of the descriptive statistics for the significant acoustic/prosodic and conversational time-related features for the three interactions and for the comparison human/agents 108

Table 5.12 Summary of the descriptive statistics for the frequency-related features for the three interactions and for the comparison human/agents. 110

Table 5.13 Summary of the descriptive statistics for the acoustic features for the three interactions and for the comparison human/agents.. 111

Table 5.14 Differences in the strength of Emotions per conditions H-NA, H-VH and H-H..... 115

Table 5.15 Differences in the human perception towards the agents between the conditions H-NA, H-VH and H-H 117

Table 5.16 Pearson Correlation Coefficient r between body joints, voice frequency and brain states, for H-H and H-NA interactions. 118

Table 5.17 Pearson Correlation Coefficient r between upper body joints and muscles, for all interactions 119

Table 5.18 Average scores for the NARS questionnaire..... 136

LIST OF ABBREVIATIONS

AI: Artificial Intelligence

EI: Emotional Intelligence

HCI: Human – Computer Interaction

HRI: Human – Robot Interaction

HHI: Human – Human Interaction

H-NH: Human – NonHuman

VR: Virtual Reality

VE: Virtual Environment

VRI: Virtual Imanginary Environment

VH: Virtual Human

UVH: Uncanny Valley Hypothesis

CASA: Computers are Social Actors

NVC: NonVerbal Communication

SSP: Social Signal Processing

LMA : Laban Movement Analysis

NARS : Negative Attitude Toward Robots Scale

EMG: Electromyography

EEG: Electroencephalography

EDA: Electrodermal Activity

HR: Heart Rate

ROIs: Regions Of Interest

Fp: Prefrontal area

F: Frontal area

C: Central area

P: Parietal area

O: Occipital area

ODWT: Orthogonal Discrete Wavelet Transform

Fmean: Mean Frequency

RMS: Root Mean Square

LDA: Linear Discriminant Analysis

SVM: Support Vector Machines

PCA: Principal Component Analysis

ICA: Independent Component Analysis

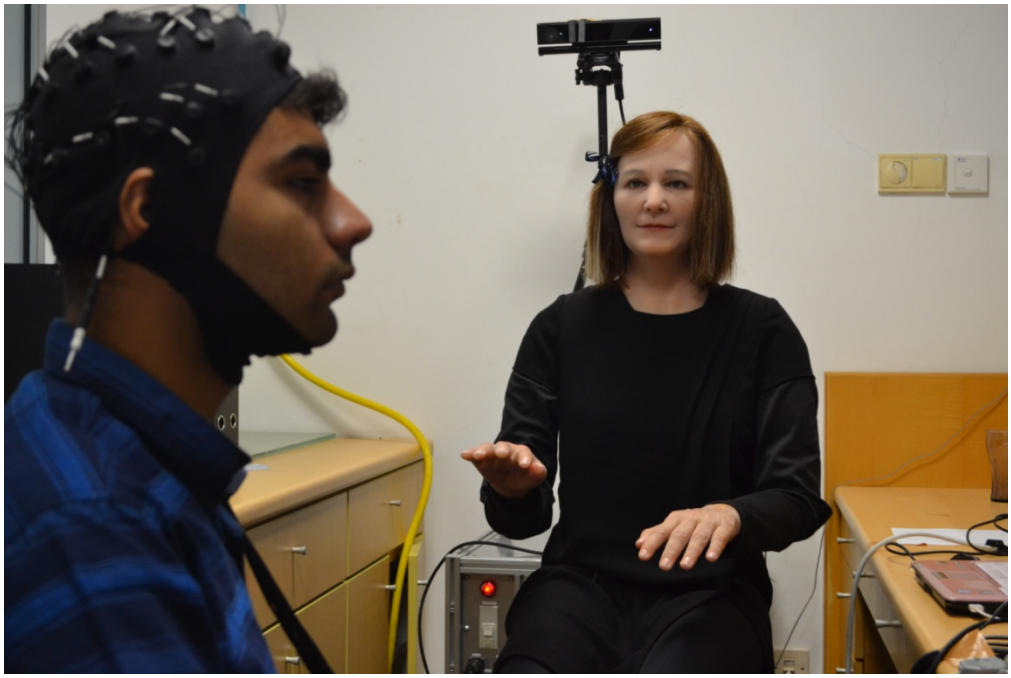
SVD: Singular Value Decomposition

KNN: K-Nearest Neighbors

CV: Cross Validation

CHAPTER 1

INTRODUCTION



"Robots touch something deeply human within us. For me, robots are all about people."

– *Cynthia Breazeal, roboticist and entrepreneur*

Introduction

1.1. Human-Nonhuman Interaction

Alan Turing's famous question "Can machines think?" inspired several researchers to start examining the potentials of Human-Computer Interaction (HCI) and later, of Human-Robot Interaction (HRI), leading to a point where technology has started to be actively involved in the communication process. However, the communication between human beings has been guided and facilitated by other factors, like emotions, empathy, emotional intelligence, and awareness. Emotions, as an inherent internal procedure, are the mirror of what we feel and they allow us to perceive and understand our environment and even ourselves. They have been classified into three major categories: the basic emotions, such as happy, sad, anger, fear, surprise, emotions based on motivation like thirst, hunger, pain, mood, and emotions based on self-consciousness and social interactions, such as shame, dignity or guilt [1]. Thus, the first questions raised here are if and how we can be based on the extraction of the human emotional and behavioral information to contribute to the design, Artificial Intelligence (AI), and Emotional Intelligence (EI) of digital humans and robots.

The importance of emotions in human communication has been supported since 1973 by Ekman [2], influenced by Darwin's work. In around 1870, Darwin was the first to examine expressed emotion by other species, like animals, trying to find similarities between them and humans and to understand if these emotions are inherited or acquired. Ekman supported that humans experience six different emotions: happiness, sadness, anger, fear, disgust, and surprise. However, research on that has even started before Darwin, with the philosophical studies of the ancient Greeks and Romans. Based on the Stoics, Cicero identified and organized emotions in four categories: metus (fear), aegritudo (pain), libido (lust) and Laetitia (pleasure) [3].

People, intentionally or not, use both facial expressions and body movements to convey emotional states and intentions to others. This conveyance is crucial for a flowing, mild social integration as everyday life is based on social interactions that demand an optimal adaptation to every context. The multidisciplinary area of computer graphics, neuroscience, psychology, and artificial intelligence is in its golden age of trying to explore, decipher, decode and model such human emotional behaviors and expressions to enhance the field of social interactions between humans and between humans and technology. To decipher and interpret the features and the boundaries between humans and technology, a first step is to compare human-human interaction with the one between humans and machines. The research of human-human (H-H) communication can reveal the most useful information for enhancing the HCI and HRI fields and thus, it

can clearly be stated as a starting point. It has already been proven that people are more willing to discuss and even disclose private information when computers follow and present human-based conversational rules [4].

Baylor [5] stated the three main factors that can characterize a natural social interaction between a human and an agent. What we characterize as an agent, based on what Ferber defined, is "a physical or virtual entity that can act, perceive its environment (in a partial way) and communicate with others, is autonomous and has skills to achieve its goals and tendencies" [6]. Thus, according to Baylor's research, social interaction is portrayed by the appearance of the agent, i.e cartoon or realistic figures, the communication features, such as gestures or facial expressions and the content of the dialogue. All this research has been based on Bandura's first theoretical social cognitive learning theory, where he supports that people learn behaviors and norms by imitating other people who react in the same way. Trying to boost this imitation, researchers are trying to create more realistic digital humans and robots to facilitate the human-nonhuman (H-NH) interaction. This realism is based on human responses and reactions as well as on human appearance. It is undoubtably that robots, compared to digital humans, can present more human features that allow them to better incorporate human behaviors. However, the question raised here is to what extent an agent needs to adopt human behaviors and appearance, known as anthropomorphism? In other words, it is required to explore when and if the humanization of robots, or in general of social agents, needs to stop and which human need each level can meet.

1.1.1 Research context and motivations

Fast-forward to the 21st century, the focus has shifted towards the relationship between humans and machines, creating the broad area of affective computing, examining emotions and perception evoked by HCI or the use of emotions for the emotional intelligence of the machines. The ultimate purpose is the facilitation of smooth communication between humans and computer-generated characters or robots.

Affective signals

The first step towards this purpose is recognizing and decoding affective signals derived from our face or body under several social interaction contexts. Around 1973, as we mentioned before, studies started examining behavioral and emotional intentions, proving that they can be predicted through facial expressions [2]. It has been shown that observers tend to activate similar facial muscle activity with the speakers' intended facial expressions [7]. Based on that, studies until 2009 have mostly examined signals extracted from facial expressions and voice [8]. The majority of such studies have used a congruent direction of gaze and body using an eye-tracking system for the gaze, EMG for the face muscles, and a

questionnaire for the self-assessment [9]. More recently, studies have started to include the analysis of the whole body and face, as they support that both play an essential role in human communication. There are even studies supporting that body movement can convey emotional information more efficiently than face under specific circumstances [10], as the body can have some advantages. First of all, due to its bigger size, it can facilitate emotional communication by making expressions visible at greater distances [8]. Moreover, the expressions can be noticed no matter what the position of the body is, compared to the face which needs always to be towards the interlocutor [10]. However, recent investigations have shown that the processing of body expression is similar to the one of facial expressions [11] as they both work as channels through which we convey emotions.

Robots and digital humans can improve the accessibility of various contexts. Robots and other intelligent systems are able to improve the quality of human life by providing assistance in intensive and difficult situations or even independence in the way of living for people who have the need, like the elderly or people with motor/cognitive disabilities. Nowadays, agents have the ability to embody and fill social roles [12][13][14][15]. An embodied agent can be a physical robot or a virtual character that has an identifiable body and can use modalities like voice, gestures, or facial expressions to communicate. The main differences between a virtual and a robotic agent are the physiology of the human face, the natural neck motion, the shared gaze but mostly the physical presence [16]. For children, adults or the elderly, agents can play a role of service, but the question is what level of effectiveness. We should not forget that a nonhuman agent cannot substitute a human, no matter its appearance but can approach the reactions and behavior of the latter.

Nature and design

Although the continuous effort of the existing studies to enhance the domain of HCI and HRI by addressing all the aforementioned features, it seems that the fluidity and the naturalness of the interaction have not been completely achieved yet. This has as a consequence to people who still prefer human communication in any context. In a study, for example, robotic and digital agents were compared through a video setting in an educational context, as instructors [16]. It was found that attitude was more positive towards humans compared to robots, but agents have the potential to act as an alternative with the strict requirement that they are designed well. Moreover, humans tend to be more open, self-disclosing, outgoing and in general more positive when interacting with another human compared to an AI agent [17]. The same was verified by another study where humans when talking to another human instead of a computer, tended to be more talkative and spend more time in the conversation [18]. This preference can also be an outcome of the low

degree of naturalism. It has been found that people laughed when they had to respond to a robot's greetings, admitting that they found the movement unusual and foreign during their interaction [19].

However, there are a lot the studies supporting that the design of agents should not completely rely on humans' nature. We need to find the key point where humans socially approve and accept social agents and social agents have as a principle the human needs. The definition of how humans evaluate robots or any type of social agents, as well as the documentation and interpretation of humans' reactions towards them, can help us identify the weakness of the up-to-date technology and improve its functionality and design.

Anthropomorphism can be derived through several characteristics, like the external design, motion features, communication skills, hypothetical emotions, and the sense of autonomy [20]. On top of all these, the human's expectations and imagination take place to guide how the aforementioned elements will be interpreted and understood. Other secondary factors that can influence the perception of anthropomorphism are gender, culture and language. Another important factor to be considered is the group dynamics and how these can affect the quality and the balance of conversation [21]. The role of human imagination during HRI in the perception of human-likeness and the abilities of the robot has already been examined by MIRALab before [20]. A realistic humanoid robotic head called Eva was used, with a high ability to imitate human facial expressions and two different situations were tested: a human interacting with a robot and a human observing an interaction.

Figure 1.1 The humanoid robotic head Eva used by Zawieska et al. at the University of Geneva to assess how the role of human imagination in the perception of human likeness and intelligence of the robot [20].

The results were very interesting and very promising for future research. The most important was that the observers found the robot more anthropomorphic and they tended to attribute more abilities, like intelligence, compared to the ones who interacted with Eva. Thus, they concluded that the less the interaction occurs in a controlled environment, the higher the human tendency to associate human characteristics to the machine. This outcome can raise again a question: At what extent does a robot, or in general a social agent, need to be humanlike? In the end, is it subjective? However, it is worth mentioning

that the reasons, stated by the participants who interacted with the robot, for perceiving it as less human-like were mainly the robot design (close distance and physically present interaction so limitations of the appearance were obvious) and the context of the scenario.

Robotic head Eva has been used by MIRALab in several types of research to explore the human-robot relationship. In [22], memory and emotional aspects of the robot in a long-term interaction experience have been studied, highlighting the role of memory in task engagement. The importance of memory and emotion in H-NH interaction has also been investigated and reviewed by [23, 24]. Eva has also been used to enhance emotional decision making by integrating attachment and learning to it, in a study towards a more empathic nonhuman agent [25]. MIRALab has also provided some insight regarding the H-H audio and visual social behavior measuring sociometrics [26].

However, to characterize a social interaction between a human and an agent successfully, two main points need to be fulfilled. First of all, the human, while interacting, needs to produce social signals and expects responses that will give a continuation to the communication. On the other hand, the agent should receive actively the human feedback but also respond in a way that can maintain the interaction. However, there are two main challenges regarding the reaction of the agent; the cultural context and the followed rules [27]. To wit, robots or digital humans are not yet trained to get adapted to any kind of interlocutor they face as they usually obey some predefined rules. Even people have difficulty coping with cultural settings they are not familiar with, thus it is normal that artificial agents are not yet at the point where they can distinguish the different contexts and adapt their behavior to them. However, the big challenge is how this problem can meet its solution. And this goes back to our first speculation on how and if we can enhance the AI and EI of an agent. The second challenge regards the internal rules that agents need to follow to execute an order. By order, we mean any kind of social cue an agent will generate to fit in the communication framework. Especially, if everything is predefined, and this predefinition consists of limited yet resources, the naturalism and the reaction of the agent lack depth. Indeed, it is not easy to decipher the complexity and the richness of real-life social communication but this is the key engaging humans during a H-NH interaction more efficiently and for a longer period.

In summary, H-NH interaction, including affecting computing and social signal processing, consists of multiple interdisciplinary research features that finally motivate us to conduct our research:

- Body motion analysis (motion and gestures features and dynamic)
- Face expression analysis
- Voice recognition and analysis (audio features for speech-based emotion analysis)
- Physiological data extraction and analysis
- Psychometric parameters

- Comparison and correlations of modalities (machine learning techniques)
- Several types of social interactions (with other humans, avatars, or robots)

1.2 Objectives and Contributions

Our basic research efforts are confined to comparing the communication between humans with the one between humans and technology, concluding to a disclosure of the humans' needs towards H-NH communication and a suggestion of potential improvements. This objective is twofold: a) to define the difference in comparison with the human-nonhuman interaction and b) to use the collected data, extracted from all kinds of interactions, to provide behavioral, physiological, and psychological patterns and decipher the dynamics in a natural human interaction. This could create new guidelines for a better engagement between humans and agents. Our goal is to assess how humans react and perceive social nonhuman agents and their technology but also, how the latter can affect humans. The understanding of human responses can help us decipher the human needs and understand how and if we need to change the up-to-date technology and design of social agents.

Emotional information can be transmitted through verbal (speech and semantic content of a message) and non-verbal (facial and maybe vocal expressions, gestures) communicative tools that can be influenced by several internal or external conditions, like the mood of the person or the environment [28]. In this kind of research, a multimodal approach is preferred, avoiding the limitations each modality may have. Facial expressions, for example, are dependent upon context and the character and thus, they vary across cultures [28]. Speech can complement the above, making the distinction between verbal and non-verbal signals more delicate, although it can also be limited if it depends on the language [29]. However, it is usually accompanied by gestures that people usually do unintentionally when interacting with each other, like hand and posture movements or gaze. When a unimodal modality has been used, then it concerns physiological signals, and usually the use of EEG.

However, in a social context, specific emotions can be developed facilitating communication by enhancing the trust and belief between the people. These emotions are called social or moral emotions, like shame, empathy, jealousy, or admiration and they are dependent on the behaviors, feelings, and actions of other people [30]. Most of the conducted research on affective computing has been focused on basic emotions and dimensional models. Specifically, emotions, to be examined and categorized, need to be defined in a dimensional space, with the most commonly used the one of Arousal/Valence, as it forms the primary orthogonal dimension of the affective experience, or they can be examined in discrete states like happiness and sadness [31]. However, social interaction is a complex situation, consisting of emotions, attitudes,

cognitions, personality traits that it is not easy to be deciphered and consequently translated to the machines. It is said though that attitudes are modified through experience [32] and emotions are shaped based on previous experiences a human may have with other humans or other kinds of technology [21]. There are several studies though that have tried to create a link between humans' reactions and emotions, based on voice [33, 34], body [35], or brain [36–39], facilitating the creation of behavioral patterns

The current research, with its multidisciplinary approach, aims to address possible limitations of the existing technology used in the broad area of human-robot interaction up to now. Although different kinds of agents have been used to contribute to several domains, like education, health, entertainment, both in virtual and physical environments, the digital human or robot that will make a human feel comfortable interacting with another human has not been reported yet. What is mainly missing from the up-to-date state-of-the-art is the direct comparison of any kind of nonhuman interaction with the original human-human communication and the clarification of human behavioral patterns in the context of both physical and technological frameworks. What we need to do is to keep studying human-human communication, and not only features of the HCI or HRI, as what we lack is how we, as humans, react in several contexts of communication. The extraction of human features in such contexts, as vocal features, gestures, or body movements, and physiological features like brain or muscle signals can complement the existing technologies and studies. Moreover, the way humans react to computer-mediated characters and virtual environments can be a tool to decipher and understand existing human communication theories that can also support the aforementioned. Towards this direction, "robotic psychology", aims to find and cover the gap between humans and robots by shading some light in features peculiar to HRI and consequently, more broadly, to HCI [21]. It is important to orient the research towards humans extracting human features, revealing humans' needs and integrating them into the technology. The latter can ensure a more successful and efficient collaboration between humans and robots or other kinds of similar technology (i.e. avatars).

To sum up, our main contributions are the following:

- A database of human reactions extracted from H-H and H-NH interactions. It is a collection of voice, brain, muscles, and motion of forty participants, including 29 men and 11 women, aged 21 to 65 years old.
- A complete analysis of human behavior and perception in different technological interactive experiences (virtual reality, digital humans, and robots)
- A machine learning model that can differentiate human voice while speaking to a nonhuman agent and to another human.

Thus, through this kind of research and by creating patterns for the human nonverbal communication, we can contribute to the enhancement of naturalism of every kind of agent, offering a higher level of understanding in the context of everyday communication. To the best of our knowledge, such multidisciplinary and in-depth documentation, analysis and comparison between H-H and H-NH interactions have not yet appeared in the current literature, as illustrated in Chapter 2.

1.2.1 Limitations

In the following list, we provide a series of limitations, derived from the current literature, that assisted us in driving our research:

- The *unnatural scenarios and tasks and the intervention of the researchers*. This limitation has also been noticed in a very recent study conducted by Stock et al., highlighting the restrictions a laboratory environment may provoke to the intensity of an interaction [21]. According to the need of each experiment, some tasks don't necessarily correspond to real-life scenarios and consequently, they create biases and affect the outcome. An example of such a task can be the study of facial emotional expressions, as emotions in real life are represented by the whole body and the restriction of it can unintentionally cause an issue. Moreover, there is in general a gap between the interaction scenarios that people face in a laboratory environment by having to actively participate or passively observe a social interaction. It has been proven that the scene of the interaction can influence the response of the persons involved and their perception of emotional expressions. This effect is called the perceptual bias effect and it was thoroughly studied by Van de Stock et al. [40]
- *Most of the studies examining human-computer or human-robot interaction have been based on the observation of images or video without providing a real physical interaction* that can conceal a lot of important information. However, the interaction may differ if it concerns a humanoid or a robot with a mechanical appearance and it is true that there is a limited number of studies that have used humanoids with realistic physical appearance. The lack of this physical interaction deprives the sense of physical presence and embodiment, which is the mutual influence of the physical environment and the actions taking place in it.
- *The small sample size*, which affects the reliability and the validity of the results.
- *The limited combination of modalities and extracted features.* The focus of the current literature is on the extraction of specific and limited human patterns that cannot provide a deep understanding of how and why humans react and think like that towards this kind of technology. Based on the existing studies, the most common combination is the one of facial expressions and speech. A detailed description of the type and the number of modalities used are presented in chapter 2.

- *The direct comparison between the human-human and human-nonhuman interactions to extract and compare the differences in the human body*.

1.2.2. Research questions

Based on these limitations, the current study intends to answer the following research questions:

1. **Can a simulated experience, like Virtual Reality, activate regions of the brain, affecting behavioral, motor, or other functions?**

A thorough analysis of brain activity under different conditions of physical and virtual environment have been measured. The answer to this question can also help us to identify the optimal environment for the use of digital humans.

2. **Is there a difference in human perception when interacting with a human and an identical human-like robot?**
   2.1 Is there a difference in emotions and motivation when simply interacting with a human and an identical robot?

3. **How do humans' voice and body react when interacting socially with humans compared to nonhuman agents?**
   3.1 Do gender or ethnicity affect human behavior when interacting with a nonhuman agent?

We compared physiological, motion and psychological human features under different social interactions. Although the proportion of our participants was not well balanced, we examined possible differences and similarities due to gender or ethnicity.

4. **Can changes in agents' up-to-date technology and design be related to smoother, more pleasant, and efficient human-nonhuman interactions?**
   4.1 Can different roles of a robot change the perception and preference of the participants?

Based on our results from the second research question and the testimonies of our participants, we want to reach a conclusion regarding the current nature and design of nonhuman agents and how they can affect humans' acceptance. This question, along with its subquestion, was answered via two different experiments. Lastly, we end up with some features that could possibly make an agent react physically and vocally in a more human-like way, being more accessible and accepted?

1.3 Potential Applications

Digital humans and robots have started to be applied in many fields like

- Education (i.e teaching assistance)
- Medicine (i.e rehabilitation methods, medical 3D games with virtual guidance)
- Entertainment (i.e gaming with characters)
- Psychology (i.e robots as social companions, virtual platforms for elderly home assistance)

Figure 1.2 Examples of agents that have been used, experimentally or practically, in several fields. Minutely, the first line from left to right: a. Orpheas, the alien musician used for rehabilitation of the fine movement of the hand in stroke patients [41], b. Nao robot, used as a mediator in a multi-party support group to reduce stress [42], second line from left to right: c. PARO the baby seal used as a social companion in elderly people [43], d. a virtual human used as a lecturer in a video presentation [16]

It has already started to be proven that robotic and virtual applications, under specific circumstances such as in the medical field, have better results than the conventional interventions, usually performed by other humans. As Magnenat-Thalmann and Zhang have mentioned in their work, social robots and virtual humans can provide a very broad range of services, improving humans' quality of life [44]. The analysis of H-H interaction has a lot to give, enhancing the role and the nature of different agents in these applications. The biggest need is to decode the "behavioral loop" created during a human-human interaction. Humans have the ability to adapt to the needs of their interlocutor, modifying their expressions and providing them with emotional feedback that allows the interaction to continue naturally and efficiently. This is something that

needs to be considered in such kind of research, as it is the key for an agent to go a step further. The recent development of virtual reality systems or human-like robots but in terms of appearance are not enough to ensure a more effective human-computer or human-robot interaction.

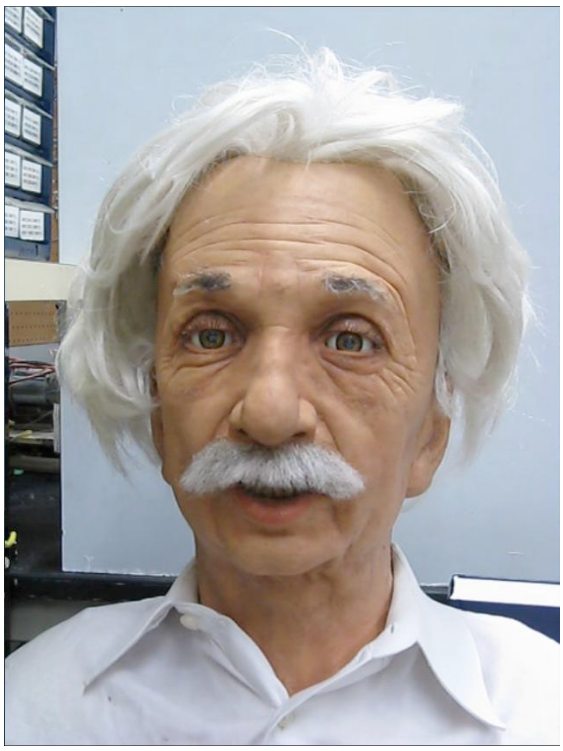
1.4 Manuscript Organization

This manuscript is organized as follows:

- Chapter 1 (*Introduction*) introduces the research, its background and objectives, and presents the research questions.
- Chapter 2 (*Related Work*) provides a detailed literature review on the domains of HCI, HRI, and HHI, focusing on nonverbal communication.
- Chapter 3 (*Interaction in Virtual and Physical Environments*) describes the methodology, research protocol, and results of the first experiment regarding the human interaction with the virtual and physical environment.
- Chapter 4 (*Human – Robot Interaction*) presents the methodology, research protocol, and results of the second experiment regarding HRI.
- Chapter 5 (*Human – nonHuman Interaction*) presents the methodology, research protocol, and results of the third and last experiment regarding the human social interaction with a physical present robot and a digital human.
- Chapter 6 (*Conclusion*) summarizes and discusses the results, answering the research questions. The conclusion is extracted from all the experiments, pointing out the significant contribution of this multidisciplinary research.

CHAPTER 2

RELATED WORK



"Computers are incredibly fast, accurate and stupid; humans are incredibly slow, inaccurate and brilliant; together they are powerful beyond imagination" . Albert Einstein, theoretical physicist

Image: © Einstein robotic head by the University of California San Diego [45]

Related Work

2.1 Introduction

In this chapter, we present a review of the relevant literature regarding our multidisciplinary subject, human-human (H-H) and human-nonhuman (H-NH) interaction. To interpret and consequently broaden the boundaries between humans and technology, the first step is to compare the human-human interaction with the one between humans and machines. The challenge here is to find a way to translate appropriately the human features to the "machine language" and to decide what features can contribute to the fluidity and the naturalness of a human-machine interaction.

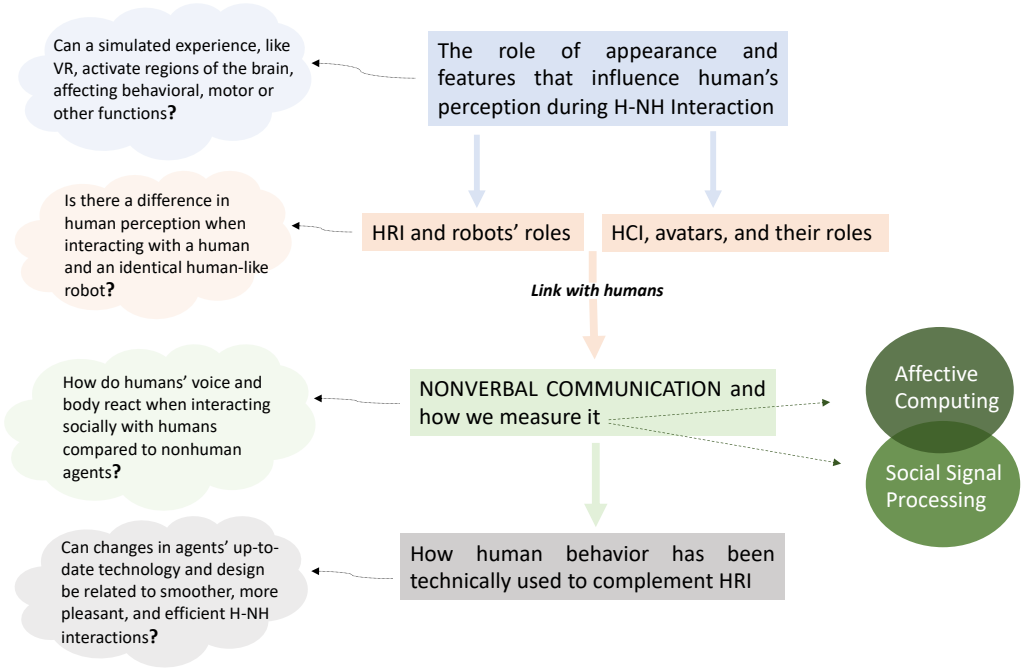


Figure 2.1 The flow of the literature review and the research questions each section addresses, as mentioned in chapter 1

The structure of this literature review is depicted in Figure 2.1. We have tried to cover all the possible aspects of such research so that we find any possible limitation that can lead us to more substantial research questions and results. Thus, we have started with the role of the human likeness as well as the role of presence and embodiment to cover the importance of the appearance and the physical or virtual presence of an agent. In the next section, we describe the basic principles of Human-Robot Interaction and we focus on social robots and their features, as we used one for our research. Section 3 describes the basis of human-avatar interaction, delimiting the definition of the term avatar and providing a summary of its use in several

fields. Section 4 creates the link between the previous two sections and the human-human interaction as it describes the principles of nonverbal communication and tries to cover the field of imitation and mimicry that seems to play an important role in human interactions. Section 5 provides us with the ways that we can measure this nonverbal communication which belongs to the broad area of affective computing and social signal processing. The basic features used for human data extraction have been noted down for modalities like audio, body movements, gestures, and physiological signals. Lastly, section 6 depicts how all these features have been analyzed and modeled, representing a human behavior that can be integrated into an avatar or a robot.

2.1.1 The role of human likeness

There are several hypotheses tested for human likeness. The most commonly used are the uncanny valley, the atypical feature, the category conflict, and the similarity hypothesis. The first one, the **uncanny valley hypothesis (UVH)**, described by Professor Mori, suggests that when a character just resembles a human, without being one, creates awkward feelings in human observers [46]. Figure 2.1 depicts the relationship between the natural resemblance and the affinity for it. The higher the human likeliness, the stronger the sensation of eeriness. There are several promoters of this hypothesis, supporting that an agent, avatar or a robot, is better to be cartoon-based rather than having a physical appearance to be more accepted by a human [5]. Research on virtual representation has proved that high levels of anthropomorphism can lead to negative effects, less trust, and discomfort [47]. Stein and Ohler supported the extension of this theory as the "uncanny valley of the mind", where they argue that it is also the "behavioral anthropomorphism" meaning the human-like behavior of the agent, that can cause negative reactions [48]. Some researchers also support that it is not only the high degree of human-like appearance that can trigger this hypothesis but also a possible mismatch between the form and the behavior [47]. However, Mori et al. expressed their doubts, proving that if the agent is designed in a way that it is hardly distinguishable from a real person, then the valence becomes positive again [46]. The morphology of an agent, aligned with the uncanny valley hypothesis, may indeed influence the perception and the behavior of a person during an interaction, but the degree depends on the task. People prefer more human-like morphology when they refer to social roles or real-time interaction for example [49]. An evaluation of UVH was conducted by Lupkowski and Gierszewska in their recent work, where they used 12 computer-rendered humanoid models to test the human perception and the UV effect [32]. For their purpose, they used a subscale of the NARS questionnaire regarding human traits. The main points of their research are that the highest comfort level was noticed for a cartoon-based character and that the belief of a person in human uniqueness can directly affect his/her attitude towards an agent; the higher the belief, the more nervous the person towards the agent.

The **atypical feature hypothesis** supports that atypical features of the stimulus may influence the perception [50]. Burleigh et al. noticed that eye size constitutes such a feature. Moreover, they found that whenever human likeness was high, eeriness was low (linear relationship) [51]. Third, the **category conflict hypothesis** [50] suggests that "*when human likeness of the stimulus is comprised of a morph between two categories, the stimuli in the middle of this scale are perceived as ambiguous, leading to a negative effect*". Yamada et al. tested also this hypothesis, concluding that the most ambiguous image reflects an increased processing time [52]. Lastly, the **similarity hypothesis** by Rosenberg-Kima et al. [53] predicts that the gender similarity (male or female) and the attractivity of an agent have a more positive effect on the motivational outcome. This hypothesis was confirmed by Shiban et al [54] who used a young female agent and an older male one to test the effects on performance and motivation in the learning process.

However, the question here is if the appearance of the agent alone can influence the perception and the performance of the user, or their behavior in combination with a contextual environment play also a role. Are there measurable benefits for the user and can we reach a level where a virtual avatar or a robot can really simulate the human behavior so that we can compare the different cases and come to a conclusion about the usability of such agents? And most importantly, is it really a need of reaching this level? To answer all these questions efficiently, more research should follow, deciphering the human-human interaction and exploring the potential of integrating its features to digital humans and robots.

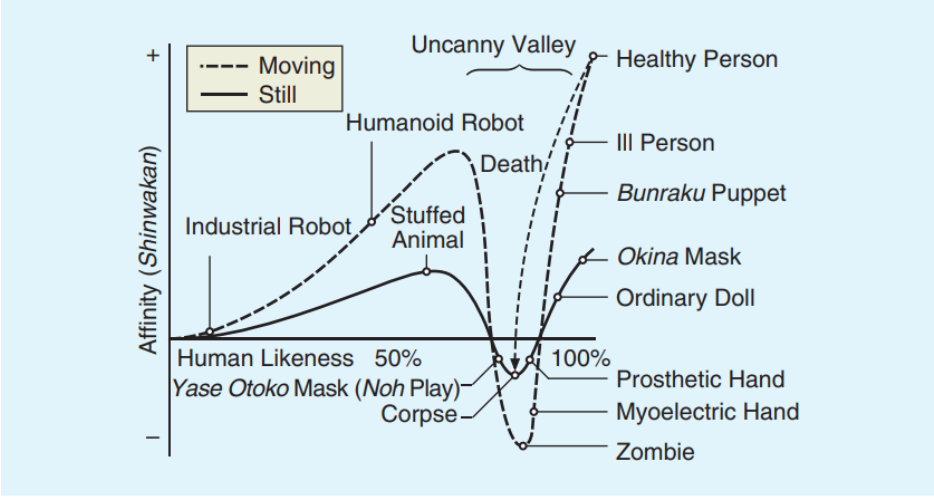


Figure 2.2 The uncanny valley as described by Mori et al. [46], depicting the relationship between the natural resemblance and the affinity for it. The dotted line represents the effect of the presence of movement.

The majority of studies, having examined the role of human likeness or the human perception towards H-NH interaction, have been evaluated through validated questionnaires. As Kätsyri et al. though mentioned, these kinds of studies cannot easily resolve the existing ambiguity in this field, so psychophysiological

studies are of need [55]. Ratajczyk et al. continued the work of Lupkowski and Gierszewska mentioned above, using electrodermal activity (EDA) and response time measurement to evaluate the UV effect and the human perception towards the same 12 characters, assessing also the role of their environment (background). Another interesting recent example is the one of Ciechanowski et al. who used facial electromyography (EMG), respirometer, electrocardiography, and EDA to examine the human-nonhuman interaction process between a human and a chatbox [56]. However, all these studies cannot imply that agents with mechanical appearance cannot positively affect humans.

2.1.2 The role of embodiment and presence

Intelligent systems have two critical features that can affect human perception during HCI: **embodiment** and **presence**. The embodiment was defined by Pfeifer and Scheier [57] as a *term which refers to the fact that "intelligence cannot merely exist in the form of an abstract algorithm but requires a physical instantiation, a body".* The level of the embodiment is dependent on the nature of the agent (physical, virtual, or even a combination of both), the morphology (i.e human-like or cartoon based), as well as the modalities it can support, and the extent to which these modalities can be carried out [58]. Other variables, like gestures, speech speed, and haptic stimuli, may also be considered as aspects of an agent's embodiment.

Whereas embodiment concerns the agent and its relationship with its environment, presence deals with the way this agent is presented to others. Milgram et al [59]defined physical and digital presence as a situation where the embodied agent can be touched, saying specifically "*whether primary world objects are viewed directly or by means of some electronic synthesis process*". Zhao categorized physical and digital presence as *copresence* and *telepresence* respectively [60]. Copresence, as a term in a sociological framework, describes the conditions under which humans interact with each other [60]. Under the umbrella of HCI and HRI, copresence refers to how the agent is displayed to the user. Zhao (2003) used two dimensions to describe the copresence. The first one refers *to "the mode of being with others"* and concerns features that can physically shape a human interaction whereas the second one refers to "*the sense of being with others"*, linked to the feeling and the subjectivity of the user [60]. We need though to differentiate physical embodiment and physical presence (copresence) as an agent, that may have physical embodiment, may not have a physical morphology presented to the user [58]. Several researchers tried to evaluate the role and the influence of presence and embodiment in virtual environments or robotics [58, 61, 62]. In the case of Virtual Reality (VR), we find these features also from a user perspective. As Slater has stated, "presence is a response to a system of a certain level of immersion" [63]. VR has some differences compared to robotics as it can create the illusion to the users that they really experience the presented situation, especially if they

are also able to see themselves in it (bodily self-consciousness) [64]. This requires that the brain cannot really distinguish any difference between the expected and the given outcome of the experience [64].

So, here we pose the first question, regarding the effect of physical presence. Do people react differently in an interaction with a copresent agent (robot) compared to a telepresent one? Research up to now has proved that psychological responses between these two situations differ due to a variety of reasons. Initially, one reason is the size of the agent and consequently the influence it can have [65]. Robots that are physically present have usually a bigger size than a virtual agent displayed on a screen. As Huang et al. have mentioned, taller individuals tend to provoke a bigger social influence [66] and thus, the larger size of the physical robot may be more imposing, having a stronger impact.

Distance is one of the main aspects of presence, as Zhao also supported [60], which can be divided into physical and electronic proximity. This leads to the second reason which is the physical distance between the user and the agent as physical proximity is normal to have different effects compared to the electronic one [67]. Moreover, the interaction with a physical agent allows a better understanding of its morphology and motion, creating a more familiar environment with the user. In general, it has been shown that physical presence can improve the user's behavior as well as increase the level of enjoyment and trust [58]. In the case of the same appearance, a recent survey showed that 79% of the up-to-date studies favored a robot that is copresent compared to a telepresent one [58].

The next question derives as a continuation of the latter and examines the effect of physical embodiment. Do people react differently interacting with a physical agent compared to a virtual one? One reason for which the embodiment may result in the psychological processing of the user is the degree of realism [65]. Han et al. compared, with the use of functional Magnetic Resonance Imaging (fMRI), real and virtual visual worlds through the observation of movie or cartoon clips, aiming to provide information on how we perceive characters in real and virtual worlds [68]. They concluded that the perception of real-world characters triggers the medial prefrontal cortex (MPFC) of the brain and the cerebellum which act as an online representation and empathy of mental states of others, whereas cartoon clips of humans and non-human agents activated the superior parietal lobes which are associated with attention when referring to actions [68]. The cartoon-based clips also engaged the occipital area of the brain which is linked with the visual attention mechanism.

Studies that have examined the influence of physical embodiment separately from the physical presence, comparing telepresent robots to virtual avatars, reported no significant results [58]. However, what if the physical embodiment and the physical presence are combined? The majority of the studies have supported that people prefer the physical presence of a robot to a virtual avatar [58], having also significant effects in several behavioral responses like performance, attention [69], and response speed [70]. However, gesturing

has been proved to play an important role in the response of people during HCI. Thus, to complement the above, people prefer copresent agents, compared to telepresent robots or virtual agents, but only when they use gestures to complete their interaction [71].

In general, Jamy Li proved through his survey that physical presence plays a greater role in psychological responses to an agent than physical embodiment [58]. So, it seems that no matter the nature of the embodiment (virtual or physical) which constitutes a feature of the character, the presence is the one that can directly influence the response and behavior of the people [58]. To wit, what matters is how the agent will be presented to the user and finally, how the embodiment can allow that. However, there is a limitation in this field as there are not a lot of studies that have used avatars of high-level naturalism, decreasing the effect of human appearance.

2.1.3 Other features that can influence human's perception during HCI

Eye gaze and facial expressions

Another important feature that has been tested in such interactions is the role of the eye gaze. Eye gaze is one of the most important features of human behavior while a social interaction as it can serve several purposes and functions like enhancing attention, revealing emotional information, preserving engagement. Therefore, it has been proved that the physical presence plays a greater role in the gaze's perception compared to physical embodiment and thus a robot's eye gaze can be more accurate than the one of a virtual agent [62].

Studies that have examined and proved that through facial expressions the behavioral and emotional intentions of another person can be predicted, started around 1973 [2]. It has been shown that observers tend to activate similar facial muscle activity with the speakers' intended facial expressions [7]. This reaction has been characterized as Rapid Facial Reaction (RFS) [72] constituting an affective reaction that can occur automatically after the stimulus presentation. The majority of such studies have used a congruent direction of gaze and body using an eye-tracking system for the gaze, EMG for the face muscles, and a questionnaire for the self-assessment [9]. However, the importance of the body's direction started to raise questions and more recent studies [73, 74] examined the influence of the difference in body and gaze direction. Thus, although it has been shown that the gaze is one of the major indicators of socio-communicative dimensions, it has been finally proved that only when it is combined and congruent with the body orientation, it can modulate emotional experience and attention.

Humans can express different kinds of emotions while interacting with different types of agents, under the same circumstances. A priori, the communication between human beings has been guided and facilitated

by the existence of emotions. Emotions, as an inherent internal procedure, are the mirror of what we feel, allowing us to perceive and understand our environment, including ourselves. It has been proved that people experience more positive emotions when interacting with a virtual agent that provides positive feedback instead of a negative one [75, 76]. Mollahosseini et al. (2018) studied the perception of people towards facial expressions of a virtual agent, a copresent retro-projected robot, a telepresent robot, and a video recording of a human and they found that the emotion recognition rates differentiated among the several agent conditions. In other words, humans perceived, and consequently expressed, differently the emotions based on the nature of the agent [62]. They proved that the physical presence plays a greater role in the gaze's perception and thus a robot's eye gaze can be more accurate than the one of a virtual agent. Lazzeri et al. (2015) also proved that emotions that are expressed through facial expressions, can be better perceived on a robotic agent than a virtual one [77]. On the contrary, virtual agents seem to be more effective when it concerns visual speech due to the computer graphics that can provide a better accuracy on the realism and the animations [62]. Kompatsiari et al. in their recent study, examined human-robot social interaction, measuring EEG signals under several gaze cueing, proving that a humanoid robot with mechanistic eyes and human-like characteristics can activate similar brain attention mechanisms as another human would do [78]. Gaze coordination is a crucial part of a well-designed nonhuman agent.

Voice and gender

Another variable that can influence human' perception during a H-NH interaction is the voice of the agent. The first to examine the role of the voice nature (human-like, robot-like), as well as the gender of the agent, was the Eyssel et al. They stated that idiosyncratic features of both users and robots play a crucial role in the acceptance and the user-friendliness of the technological system [79]. They also indicated that same-gender robots were perceived more positively. Towards this direction, more studies have studied the importance of congruence between the visual appearance and the voice characteristics of robots in humans' perception and expectations [80, 81]. The more human-like the voice, the higher the expectation of anthropomorphism. Thus, it is not clear yet if there is always a need for nonhuman agents who mimic the human voice. Trying to answer this question, Google developed an artificial intelligence technology, called Duplex, which imitated completely the human voice, including pauses and hesitations but without a body [82]. The result was the rise of several ethical questions regarding how confusing and deceiving can be for a human to vocally interact with such a technological interlocutor.

However, it has been proved that nonhuman agents with vocal entrainment, able to change features like pitch, speaking rate, and intensity to mimic the user, have a positive outcome in the perception and trustworthiness [83, 84]. The general outcome of all these studies though is that the key to increase the

acceptance the trust that can enhance the flow of a social interaction is to match as much as possible the user preferences with their expectations [85]. In other words, if we want to have nonhuman agents with a natural human-like voice, their appearance should accompany the latter. However, it has been shown that an agent can have a positive effect on a user only when the voice it supports is human-like and not a machine voice [86].

Once again, all the aforementioned studies have validated their results through questionnaires. For a more profound conclusion, more research is required and physiological features need to be measured.

2.2 Human-Robot Interaction (HRI)

There has been a lot of research trying to decipher human behavior and perception when interacting with robots compared to other humans. There are even movies that describe such interactions and, even if we consider them as science fiction films, we are at a point where people have started communicating and meeting social robots in a real-life context incorporating personal or professional roles [49].

It has already been shown that the first reaction of people toward an initial communication with a social robot is a feeling of uncertainty and decreased anticipation [49]. However, Edwards et al. suggested that this behavior is a result of the deviating social communication pattern, that leads to the alteration of the "script" and expectations during a human-human interaction. Humans, unintentionally, follow a script when interacting with each other, adapted to various social situations. One of the roles of HRI research though is to decode these scripts and allow similar behaviors to take place during a human-robot interaction.

Communication has been described by Kellerman (1992) as a "heavily-scripted procedure" [87]. In the framework of this procedure, humans are used to interacting with other humans, creating an anthropocentric expectancy in communication. However, despite these expectations, it has been supported that people tend to treat computers or other social intelligent technology as if they were people, by applying similar social scripts like the ones used during a human-human interaction. Reeves and Nass (1996) first illustrated this opinion with their Computers Are Social Actors (CASA) paradigm, showing that people mindlessly relate to machines and apply social rules as if they were indeed real people, even if they are aware of their incapability to embody emotions and intentions [88]. Reeves and Nass, in the same study, also suggested that people treat televisions like real people. This was confirmed by Nass and Moon (2000), who examined users' responses to different kinds of televisions and they concluded that humans perceive them also as social actors [4]. In general, Nass and Moon supported that people tend to focus on the social cues, even if they are a few, bypassing the asocial features of the entities. CASA has been already involved in several studies in a broader field of research including AI and social robots. More recently, Yi Mou & Kun Xu

compared the initial human-AI social interaction with the one between humans in terms of personality traits and communication attributes [17]. They support that their outcome complements the CASA paradigm as they found that people can change their behavior towards social actors if they are aware that they will interact with an AI.

Edwards et al. showed that the human-like morphology can satisfy this anthropocentric expectancy during an interaction. They also confirmed the hyperpersonal model, launched by Walther, based on which computer-mediated communication can sometimes surpass a face-to-face interaction in terms of intimacy and liking [89]. Thus, they concluded that according to the context of the discussion an interaction with a robot can increase the level of attribution of social presence and decrease the degree of uncertainty.

While the boundary between human-computer and human-human interaction is described by the CASA concept, social psychologists maintain doubts regarding the psychological invariance that can characterize a person across several different situations. This is the so-called personality paradox or consistency paradox, describing that a person can present different personality traits and behaviors under different circumstances. Attempting to solve this paradox, in the framework of human-computer interaction, Mischel and Shoda developed the Cognitive – Affective Processing System (CAPS) [90]. Mischel wanted the psychologists to think like mechanics and valuate people's responses according to particular conditions. According to this model, the personality system encompasses mental representations consisting of various cognitive-affective units (CAUs) that include a person's goals, beliefs, values, affective responses, and memories [91]. Different CAUs can be activated under different conditions and different contexts, shaping accordingly the behavior of the individual. Consequently, when interacting with a machine, some people may feel more confident during the interacting process whereas others can feel confused and frightened. Therefore, based on the CAPS model, when interacting with a machine, humans' behavior and reaction should be different than the one presented when communicating with another human [17].

However, it has been shown that putting robots in an anthropomorphic framework, by giving to them a personal name and even a story to follow, can affect human behavior and reaction towards them [49]. Moreover, visual gender-stereotypical cues can also affect the perceived robot's gender [92].

2.2.1 Social robots and their features

The idea of robots, as a mechanical agent serving specific purposes, has started a very long time ago, described even in Greek mythology. However, robots with natural language features, able to participate in a conversation, appeared in the 1990s, with the example of MAIA [93] and RHINO [94]. These kinds of robots were developed to cover a specific range of applications and consequently had some limitations, like

the limited non-verbal communication, the difficulty in the perception of human speech, the specific pre-defined range of responses [95]. All these restraints of the '90s have become the inspiration of the next years' research trying to understand and enhance the features of human-machine interaction.

Robots have been tested in several roles serving various applications where verbal and nonverbal communication are needed, like assistance and companionship [96, 97], receptionist [98], educational purposes [16, 99], museum robots and tour guides [100], or even involved in art, like musicians [101]and dancers [102]. In all the above applications, the main goal is the fluidity and the naturalness in the communication between the human and the machine, for any verbal or nonverbal feature. To succeed in this, researchers had to address limitations like breaking the "simple command only" barrier, coordination of motion and nonverbal communication, affective interaction, multiple speech acts, mixed-initiative dialogue, etc [95]. On the contrary, this kind of restraints has already been addressed in the virtual world since the early seventies, with Winograd's SHRDLU program that could support different speech acts and basic mixed-initiative dialogue [103]. Due to the lack of the physical entity of a robot, VR was easier to be developed faster and in a different way than the area of robotics. We can assume that this is why people are more used to this technology, expressing also a higher preference towards it. Nevertheless, the main difference between robots and virtual agents is the physical embodiment.

Birmingham et al. examined a new role for robots, as a mediator in a fMR multi-party support group [42]. The role of the robot was to motivate people to speak to each other and overcome their stress by increasing their sense of trust. Participants however declared at the end that the robot made the discussion mechanical, with a lack of real flow and they noticed the specific features of the robot responsible for that. As the authors used a Nao robot, the participants noticed the lack of humanity first of all in the expressions of its face. Thus, in line with other studies, facial expressions play a crucial role in efficient interactions. For example, Zawieska et al. highlighted the importance of facial expressions as the majority of their participants attributed the intelligent behavior of the robot used for their experiment to its facial expressions [20]. Moreover, Birmingham et al. found that the sound of the robot was not natural and consequently, non-native speakers had difficulty understanding its voice [42].

One of the most important items that has been addressed by both worlds (robots and virtual agents), is the affective/emotional aspect. Affection during human interaction plays a crucial role as it is directly associated with learning processes, persuasion, and empathy [95]. Pioneering work in this domain was made on virtual avatars like Steve [104] or Greta [105] that became the inspiration for Cynthia Breazeal to develop the Kismet robot, an expressive mechanomorphic robot head with perceptual and motor modalities that can support multiple facial expressions [106, 107].

The second most important feature is the one of motor and nonverbal communication coordination. People, when interacting with each other, they use several kinds of motor actions head nods, hand gestures, gaze movements, and of course lip-syncing [95]. There has been also stated that humans use lip information to perceive better a communication, the so-called McGurk effect [95]. Thus, to support even the basic level of naturalness during an interaction, agents should be able to use some of these features to accompany their sound production.

Social robotics is a rapidly increasing field aiming to develop robots capable of socio-emotionally interacting and communicating with humans serving several domains like education, health, entertainment [62]. The research and recent technologies are trying to define the best choice between robots and virtual agents, best suited for the needs of social interaction.

Table 2.1 Examples of social robots from several domains in a chronological order

Robot's name	*Reference*	Year	Type	Role
WABOT [108]	*Sugano and Kato*	1987	Humanoid	Piano player
PARO [109]	*Shibata et al.*	1997	Baby seal	Social reintegration of elderly people
Care-O-bot [110]	*Graf et al.*	2004	Non-humanoid	Home assistance for elderly people
RI-MAN [111]	*Odasima et al.*	2006	Humanoid	On-site caregiver / lifting humans
ROBOTA [112]	*Billard et al*	2007	Humanoid	Robot - assisted therapy for autistic children
IROMEC [113]	*Marti et al.*	2009	Non-humanoid	Children companion for knowledge enhancement
KASPAR [114]	*Dautenhahn et al.*	2009	Humanoid	Robot - assisted therapy for autistic children
SHIMON [115]	*Hoffmann and Weinberg*	2010	Humanoid	Playing of percussive instruments
NADINE [116]	*Kokoro and Thalmann*	2013	Humanoid	Social companion
SOPHIA [117]	*Hanson Robotics*	2016	Humanoid	Social robot
LIO [118]	*Miseikis et al.*	2019	Mechanical	Personal care assistant tasks
TENGAI [119]	*TNG*	2020	Robotic head	Job Interviewer

QTRobot V2 [120]	*LuXAI*	2020	Mechanical	For AI research and Teaching
ERICA [13]	*Inoue et al.*	2021	Humanoid	Social companion / Job interviewer
AMECA [121]	*Engineered Arts*	2021	Humanoid	Social Robot

Although the continuous effort of the existing studies to enhance the domain of HRI, it seems that the fluidity and the naturalness of the interaction have not been completely achieved yet. Consequently, people still tend to prefer human communication in any context. Jamy Li et al. compared videos of robots and virtual agents acting as instructors in an educational content, concluding that attitude was more positive towards humans compared to robots [58]. However, they noticed that the agents could potentially take this role but only if they are designed well. In this direction, Yi Mou and Kun Xu, as well as Shechtman and Horowitz, found that people tend to be more talkative, outgoing, spend more time in the conversation, and present more self-disclosure when interacting with another human than with an AI agent [17, 18]. The first question raised here is if this choice is an outcome of a low degree of naturalism and a lack of expressivity. Fischer et al. for example found that people who had to respond in a robot's greetings felt weird and laughed, reporting the unusual nature of the movement [19]. Thus, can this preference be affected by the nature of the agent? Mollahosseini et al. studied the perception of people towards facial expressions of several agents in a video recording, concluding that each agent's condition affects the emotion recognition rates. To wit, participants perceived and expressed their emotions differently according to the nature of each agent [62]. Lazzeri et al. also compared facial expressions between a robotic agent and a virtual one, proving that emotions through facial expressions can be better perceived when presented by the former [77]. However, Tsiourti et al. recently showed that human perception and believability towards robots is also affected by the accordance of this reaction with the overall context of the interaction [122].

2.2.2 Robots and their roles

Robots have been tested under different roles, with the most common to be interviewers, teachers, customer guides, and companions.

Robots as Interviewers

The world's first robot designed to carry out unbiased job interviews by Furhart Robotics and Stockholm's KTH Royal Institute of Technology is called Tengai and it is a robotic AI head [119]. Another example of such robots is the Australian Matilda [123] and the Russian Vera [124]. Nowadays, a lot of companies in

their effort to conduct impersonal and unbiased interviews, prefer to use artificial intelligence. HireVue is an example of that, where the AI analyzes voice, body language, and facial expressions to determine the qualifications and suitabilitity of the candidate [125]. Hubert is also a similar, more recent, AI recruiting platform [126].

Robots as Teachers

Another frequently used role is the one of teacher or classroom companion. Robots and AI have been used at all levels of education from kindergarten [127], elementary school [128], high school [129, 130], to universities [16, 42, 131]. However, at least for the time being, such robots need the human intervention to prepare the proper material, although they can serve as a motivation to students [16]. Large scale projects, such as "The Robotics Alliance Project" by NASA [132] and "The telepresence Robot Kit" [133] have been used to motivate students to be involved in technological fields of study.

It has also been proved that robots can support language development, enhance writing skills and teach sign language, letting teachers give more time to other groups [134]. In general, robot teachers present pros like new ways of teaching, preparation of children for a world of AI-based products but also cons like reduction of human interaction and limitation on what robots can do [135]. Nevertheless, they have the potential to teach successfully but more research is required to identify humans' needs and preferences.

Robots as Costumer guides

Guiding customers is a suitable task for a robot. "Service robots lend a hand at China's banks and railway stations" [136] and "Will robots take your job?" [137] are a few of the increasing news headlines about the emergence of service robots. In [99], a robot was explored in possible tasks as a guide in a shopping mall. The robot interacted with the customers and provided shopping information. However, the robot was partially controlled by a human operator. In recent years, robots have also been tested in frontline service, assisting human users [138]. The social robot Nadine has already been tested as a support desk in an insurance company [139]. Some tasks with repetitive or back-breaking nature have already been replaced by robots, especially in Asian countries. Nestlé, for example, has placed hundreds of robots to sell coffee on shop floors in Japan [138]. However, what is important is to examine how a robot can guide as effective as a human and what human characteristics we need to implement in a social robot. This work was recently partially done by Heikkila et al. who studied how a social robot should be designed to give effective proper guidance in a shopping mall [140].

Robots as Companions

Although it is evident that a robotic companion could not replace human interaction, friendly and well-designed robotic creations have started the effort to fill the gap of loneliness, especially in the elderly. Table 2.1 summarizes some of the most commonly used robotic companions, starting with the furry robotic seal Paro [109], dedicated to the social reintegration of elderly people. Robots like Jibo [141] or Robota [112] have also been tried to assist children with autism spectrum disorder with success. Miko [14, 142] is another very recent example of children playmate, starting from the age of five years old. Buddy [143] is called the first emotional robot for all the family. Nadine [116] is also designed to give companionship to the elderly, supporting them [15].

2.3 Humans and Virtual Humans

Over the last years, the use of Virtual Human (VH), or Virtual Avatar, has started to be known for its effectiveness over the use of real humans, boosting users' motivation and even performance. Thus, the question of whether to implement a virtual agent or a robot is still under a lot of investigation and is considered to be completely dependent on the requirements of the task to be performed. The main advantages of a VH, as they have been stated until now, are the overall little cost of use, the easiness and flexibility of its use as it can be used anytime and from anywhere, and the dynamical anytime changes of its appearance. further possibilities can be offered like collecting and examining real-time physiological data, such as facial or movement expressions [58, 75]. For a better understanding of this comparison (Virtual vs real human), we need to mention two terms. The first one, the *agency belief*, refers to the reaction of people towards VHs and specifically to the extent to which they can believe that a VH represents a real human [76] whereas the second one, the *behavioral realism*, concerns the degree to which a VH can really behave like a real human [144].

It has been shown that different levels of agency belief and behavioral realism serve different purposes. For example, VH's low behavioral realism is considered to be suitable for interviews settings [145]. Specifically, voice-only interviews have been proved to be more effective than face-to-face ones, helping participants to feel more comfortable, speaking with a higher level of self-disclosure [146]. Moreover, participants' low agency belief seems also to be more effective in such cases [76]. On the other hand, Baylor and Kim [147] showed that a physically present agent can provoke better motivational results than a voice or a text box under learning circumstances. Several studies are supporting that embodied talking agents can enhance the engagement of the user [62]. However, Mayer and Dapra showed that an agent can have a positive effect on a user only when the voice it supports is human-like and not a machine voice [86].

2.3.1 Conceptualization and Perception of Avatars

VR Environments (VE), with their virtual characters, can offer opportunities and enable manipulations that may be difficult, or even impossible, to happen in a natural environment. In these environments, users can control, embody and interact through avatars in several contexts, shaping the field of computer-mediated communication [47]. The use of an avatar, in such kind of communication, plays a crucial role as avatars can be used as a means of influence in a variety of contexts like health communication, interpersonal communication, nonverbal communication, advertising, etc. [47]. It can also support more complex behaviors and actions, enhancing nonverbal communication through gestures or body movements.

Every avatar has each own characteristics that can include for example appearance, behaviors, or abilities and can be specified based on several factors like the users' preference and their previous experiences in such environments as well as the technological capabilities of the system. However, as Nowak and Fox. (2018) mentioned, the term "avatar" is used by many researchers without being properly defined, causing sometimes misinterpretations in the framework of the relevant studies.

The origin of the word "avatar" is derived from Hinduism and specifically from the Sanskrit word for "descent" [47]. In this concept, an avatar is the incarnation of a deity on earth, being able to experience the human aspects. Nowadays, and for more than twenty years, avatars have been acknowledged as digital representations. The term became popular mainly through the novel of Neal Stephenson (1992), who used it repeatedly to refer to characters being in digital environments [148]. Following that, a lot of researchers gave several definitions to this term trying to include features like appearance, abilities, the degree of realism, or anthropomorphism. Therefore, some definitions include terms like "cartoon-based" or two dimensional" but these are continuously evolving as the technologies advance. We often hear terms like "embodied avatar", "virtual human", "digital human", "agent". In every case, two main points are served; the avatar can represent the user in a computer-mediated environment, and it can provide the experience of interaction with the environment or with another user. The most recent definition is the one of Nowak and Fox (2018) where "*an avatar is a digital representation of a human user that facilitates interaction with other users, entities, or the environment*" [47]. They chose to use a broad definition that can be used as an umbrella, independent of any specifications or characteristics.

The characteristics of an avatar can directly influence the user's perception. For example, based on the Information Processing Theory, people can get easier affected and can pay higher attention to sources that consist of dynamism [149]. Aspects that can influence a person's perception of an agent being in a virtual environment can be technical, like anthropomorphism or realism, or in a more social context, like gender, age and ethnicity.

Minutely, anthropomorphism includes the perception of any human trait or quality such as emotions, behavior, cognition, presented in any human or non-human entity. It can be mainly increased by the image of the avatar as well as its behavior [47]. There are a lot of studies on how anthropomorphic representations can influence communication, showing that the higher level of it can lead to a more natural and persuasive [150], more attractive [151] interaction, with an increased level of social presence and engagement [152]. Furthermore, realism is the perception of how a situation or an object can be realistic, and it is often mixed up with the term anthropomorphism. In the context of realism, an avatar can be judged based on its appearance, the rendering, naturalness, and the fluidity of its movements and way of speaking. Very interesting is the work of Ciechanowski et al. who, in their recent study, used several physiological measures to examine the human-nonhuman interaction [56]. For the nonhuman part, the researchers used two types of chatbox, a simple text chatbox and an avatar. Their results concluded to the main point that a chatbox should not be designed to replace a human, not even to pretend to be one. The physiological data demonstrated that the physiological arousal is higher when participants had to interact with the nonperfect imitation of the human. Minutely, more negative emotions appeared when participants interacted with the avatar, with the general outcome that the more the chatbox was considered as inhuman or strange, the less it was preferred.

On the other hand, given that avatars are perceived as social entities based on CASA [88], there are also social factors that can influence the perception of the users. First of all, the most common categorization humans use to do is the determination of gender. As Lakoff (1987) said people tend to attribute a gender to others even when physical or biological information is not available [153] and probably this is an instinctual procedure as they believe that they can understand others or predict behaviors. Studies have proven that gender in specific contextual virtual environments plays a role in human reaction. For example, children prefer a male voice when it regards to football and a female when to princesses or make-up [154] whereas adults prefer a young female avatar compared to an older male one for educational purposes [54]. Moreover, people often try to decipher the ethnicity of a person as they believe they can predict her/his behavior [47]. A study by Eastwick and Gardner (2009), among others, showed that people were influenced by the existence of black and white people in a virtual environment [155].

Another study that proved the role of gender combined with self-similarity in a gaming environment, is the one of Lucas et al. [156] where men preferred to be represented by their avatar whereas women preferred a stranger. In this study, a photorealistic self-similar avatar was used to study the effect of the appearance of the avatar in the performance and the perception of the user under a gaming environment. Lucas et al. tried to answer the question of the importance of the self-relevance of a virtual human under a specific context and although the difference in gender, they noticed that the self-similarity provokes a bigger engagement

and connection between the user and the avatar. A similar recent study is the one of Wauck et al. [157] who used a more natural photorealistic self-similar avatar in a gaming context but with better technological features with which they respected even the gender aspect and they used different animations (male and female) for the two genders. Their results indicated that there is no difference in the performance of the user based on the appearance of the avatar and no effect on gender as well. They attribute that to the better technology they used with which they avoided any negative effect on user's experience. However, further investigation is needed under different environments and contexts to verify or contradict all these results.

2.4 Nonverbal Communication

Nonverbal communication (NVC) consists of nonverbal cues expressed by facial expressions, body movements or gestures, and voice, without the linguistic content. However, its interpretation is dependent on the intention of the perceiver as such expressions and movements can be totally subjective. For this reason, it is said that a complete understanding of NVC should include also verbal communication [158], as a holistic interaction involve both nonverbal and verbal features. NVC is studied in the framework of many applied fields like medicine and mental health, business, education. The broad area of computer science carries out extensive research on that to help the design and the development of virtual humans and robots.

The main purpose of the NVC research is the study of stereotypes in human attitudes or behaviors, or the study of emotions and their behavioral parameters. There are several functions, represented by the NVC like expressing affection, regulating interactions, managing impressions, exposing opinions, or revealing several conditions [158]. We can divide the whole procedure of NVC into three basic steps, based on Brunswik's lens model [159] given that an encoding and a decoding process are required. The first step is the translation of any emotion or a personality trait into a nonverbal cue and constitutes the encoding process. The second one is the interpretation of this translation by the perceiver and consequently, the third one concerns the accuracy of the decoding which is dependent on the available cues. Both the second and third steps form the decoding part.

2.4.1 Part 1 – Encoding process

People encode several visual, tactile, auditory, or even olfactory information from the environment revealing facts of their personal states or features of their interaction with others. This encoding is a result of a two-dimensional process where the horizontal axis covers all the static and dynamic cues whereas the

vertical one represents the consciousness during the process. Figure 4 below, as stated by Hall et al., presents the framework of recent findings of nonverbal encoding [158].

During interactions, as can also be inferred by Figure 4, static information concerning people's self is encoded unintentionally. Such information may regard the identity of the person or the gender. For the identification of the identity, we are referring to a domain of computer science called biometrics, which can include iris recognition, body odor, hand geometry, etc [160].

Other features, like voice, body gestures, or gait, have been considered to work as markers for the person's gender. Mutic et al., for example, found that women tend to be more expressive in terms of facial expressions, voice, and hands movements compared to men [161]. Moreover, it has been shown that men present a higher nervousness expressed by their legs and feet. Up to now, voice, timbre, and pitch seem to be the most used voice markers [158]. Furthermore, body and voice features, like pitch variability, gaze patterns, hand gestures, eye movements, have been used as diagnostic biomarkers for mental disorders for example, like autism, schizophrenia, or even anxiety. Wieser et al. conducted an experiment where people had to interact with a virtual avatar from a distance and they noted that socially anxious people gazed less, verifying the results of the physical environment [162]. Other nonverbal cues can be more dynamic as the person can choose their nature (intentionally or not). It has been shown that changes in nonverbal unintentional cues, as shown in quadrant 2 of Figure 2.3, can provide information about the cognitive status of the sender [158]. This has been also correlated with the existence of gestures that accompany the vocal qualities.

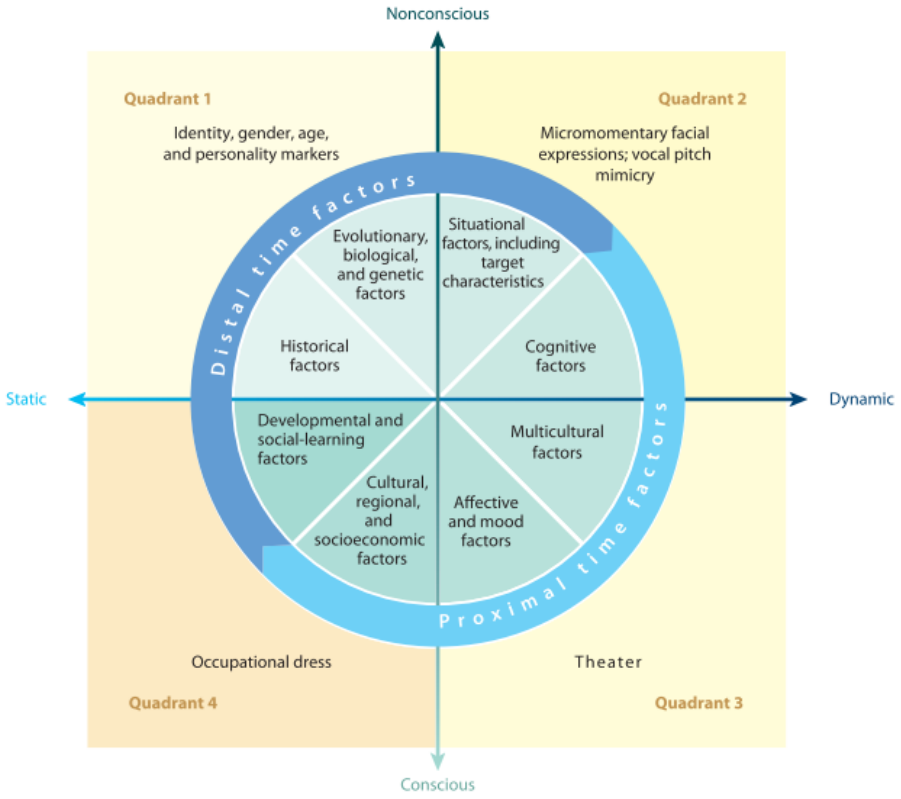


Figure 2.3 A framework of nonverbal encoding cues created by Hall et al, 2019. The horizontal axis shows that encoding can have a range of static and dynamic cues whereas the vertical axis covers the consciousness of the encoding process. Examples of encoded information are presented within the resulting quadrants [158].

All the aforementioned nonverbal cues can be both seen and heard by the perceiver. Cues that can only be heard, vocal qualities, have been examined separately as it has been shown that they can reveal features like the motivation, the status, or the interlocutor's qualities. Most used vocal factors for such purposes are timbre, pitch, speech rate, and length of pauses. The most common example is how our voice changes when we speak to babies or the elderly accordingly. The first case is called infant-directed speech and it has been associated with a higher pitch, slower speech rate, and changes in timbre [158]. These changes can be consciously on unconsciously done but it has been shown that infants can understand the difference between infant-directed and adult-directed speech by presenting higher event-related potentials (ERPs) [163]. Respectively, people change their way of speaking when addressing the elderly in a way that resembles the infant-directed speech.

Voice changes can also characterize status relations among several contexts. Under a position of negotiation, for example, people tend to use a higher but less variable pitch and to speak louder [158]. However, most importantly, vocal qualities have been correlated with affective states.

2.4.2 Part 2 – Decoding process

The decoding process is the continuation of the encoding one and it refers to the interpretation of the NVC. This interpretation is usually subjective and thus, not necessarily accurate. This accuracy is dependent on the emotional state, the personality features, the gender, the education, the motivation as well as the culture of the perceiver [158] but in any case, the cues will create a social relationship. The decoding process consists of both automatic and cognitively controlled components but it has been shown that during the first seconds or even microseconds of an interaction the first impression is created with little or no cognitive control [164]. The role of NVC has been recognized by several social-cognitive models that have tried to examine the question of accuracy in human perception. Such models started to be created around 1977 with the ecological theory of social perception by Zebrowitz and Collins [165] and they continue to be developed with a recent example the one of Zaki [166] and the model of a social cue integration framework. To measure this interpersonal accuracy, several factors are considered, like psychometric properties, affective states, and social situations.

2.4.3 Mimicry or adjustment?

Research in psychology has shown that people tend to mimic behaviors when interacting with each other. Behaviors in this sense refer to facial expressions, body postures, hand gestures, and other nonverbal or verbal cues in the context of social interaction. Mimicry theoretically can be divided into two parts: motion and emotional mimicry [167]. Motion mimicry includes behaviors that are identical in expression whereas the emotional one consists of behaviors that may not be identical but they convey the same affective state. It is also supported that motor mimicry is a part of emotional mimicry as the former can be used as a tool of expression for the latter. Psychophysiological studies examining the interdependency of behaviors in a human social interaction have proved that empathy plays a really important role as it is responsible for the adjustment of people's affective states through mimicry [168]. Mimicry is a domain that has been studied enough between human and nonhuman agents, mainly through facial EMG, examining if humans can imitate an expressive human agent, even if they are aware that the expression is not emotionally based.

Except a complete mimicry, it has been shown that people tend to adapt their behaviors when interacting with each other, according to the received emotional and social cues. Specifically, emotional expressions consist of information regarding the producer and are then used by the perceiver to adjust her\his behavior [169].

The goal is to identify mechanisms that people use to understand others, to respond and react, i.e the mechanism of imitation. Facial expressions that can convey emotions are one of the main elements of this

mechanism. Several studies have used robots as tools to examine such mechanisms of interaction as mimicry. However, mimicry is supposed to be a more complex procedure as there are two different explanations for its nature. Firstly, it can be completely motor dependent, which means that the imitative expression mirrors only the shape and maybe the dynamic of the observed one. The other explanation describes mimicry as a more internal procedure triggering emotional responses [45]. Based on the latter, opinions are supporting that mimicry cannot happen if the observer isn't convinced that the agent has mental states (psychological anthropomorphism). Hofree et al. examined the mechanism of spontaneous mimicry during a human interaction with a hyper-realistic agent with a virtual and physical presence and they tested if the perception of human-likeness can influence the existence of mimicry, using the individual Differences in Anthropomorphism Questionnaire (IDAQ) and facial EMG [45]. They found that when the agent is virtually present, mimicry occurs only when participants describe the android humanlike whereas when it is physically present participant tend to imitate its facial expressions without limitation. However, the users described the physically present android as less humanlike than the virtual one, suggesting that perception plays a role in mimicry and creating a link between the UVH and the mimicry procedure. In other words, participants tend to imitate facial expressions when the agent is physically present [17], and the direction of the body and the eye gaze can affect the experience [73].

2.5 Affective Computing and Social Signal Processing

Except the theoretical base of the NVC and the social models behind it, what is important to go beyond is the technology that can be used to enhance this field. This technology is used for the recording and the recognition of different features during several interactions of various contexts. However, for such research and high-level inferences, except the computer assistance that can automatically recognize face, body, hand, and finger movements, the existence of human observers is also essential for behavior and intention judgment. This interdisciplinary research area that aims to the detection, the labeling, and simulation of human affective states is called affective computing and it gathers researchers from various fields like computer science, cognitive sciences, psychology, and social sciences. Affective computing research has evolved from a typical unimodal analysis to more demanding complex forms of multimodal analysis. This was a natural consequence of the multimodal way with which humans communicate providing both semantic and affective information.

Emotion assessment and its technology have contributed to the improvement of interfaces that can be easier adjusted to human communication. However, the latter doesn't consist only of emotions but also of other nonverbal cues that constitute social signals and behaviors like gestures with intention or head nodding.

The analysis of these signals and their integration in the H-NH interfaces have been named social signal processing (SSP) and it is very close to the broad area of affective computing, complementing the field of emotion assessment [30]. SSP is a field introduced around 2007 by Pentland [170] and aims to complement the field of affective computing by examining social signals during H-H interaction. It differs from affective computing in two different aspects. First, it works with signals derived from groups, and second the range of extracted signals, cues, and behaviors is wider. Most of the recent studies that examine features during H-H interaction, with the aim or not to integrate the results into the field of H-NH interaction, use SSP for their analysis.

2.5.1 Affect Recognition

Affect recognition has been based on the extraction of two kinds of signals: the physiological and the non-physiological ones. Physiological signals include the Galvanic skin response (GSR), electrocardiogram (ECG), skin temperature (ST), electroencephalogram (EEG), Heart Rate (HR), pupillary diameter, or respiration patterns and the non-physiological include facial expressions, voice detection, and bodily expressions or gestures. The early works have been mainly based on the second ones and especially in visual and aural information. Ekman, and then Izard, were the first to examine facial expressions, using facial muscles to explain how the facial appearance can present an emotion [28]. Physiological signals though, in recent works, seem to have higher accuracy, with EEG being in advantage as the signal comes directly from the central nervous system, so it can provide useful and, most of the time accurate, information about internal affective states [171]. Researchers started to examine emotion recognition in 1986, using simple signals like skin conductance or heart rate [172, 173].

To increase the accuracy and the reliability of such estimations, a multimodal data fusion has recently started to be tested. In the beginning, studies started to combine non-physiological signals [171], like facial expressions, audio features, and text-based emotion recognition. In the meantime, other studies tried to use a combination of physiological signals, like EEG and eye-tracking [174]. Castellano et al. used information of four modalities, facial expressions, body movements, gestures, and speech to examine the presence of eight basic emotions [175]. Recently, Liu et al. though, were one of the first to try to combine physiological with nonphysiological signals and to integrate an emotion recognition system with the use of speech, expression, and gestures receiving also physiological signals during a human-robot interaction [175]. However, their results are not clear so we cannot come to an accurate conclusion about the combination of such modalities.

2.5.2 Affect Computation

The big question raised here is how individual affective and cognitive signals can be translated to social ones and thereby, to social cues. Affective states can act as social cues and their observation can lead to the adjustment of behavior during social interactions. For their assessment, physiological and neurophysiological signals have been mainly used, derived either from the peripheral or the central nervous system. However, such signals cannot be considered directly social signals as their use is to be analyzed to provide social cues.

Audio modality

Research on the extraction of audio features has been focused on phonetic and acoustic properties of a spoken language that have been used to train machines for emotion detection [177]. Through psychological studies, it has been found that vocal features and mainly pitch, intensity, speech rate, and voice quality can represent human emotions [3]. However, these parameters depend also on personality traits.

The most important audio features used up to now from studies on audio-based emotion analysis are:

- **Pitch** (frequency) *is the quality of a sound governed by the rate of vibrations producing it* [3]; the frequency of the sound we perceive, indicating how high or low is a tone
- **Pause duration** shows the time a person remains silent during her\his speech.
- **Intensity** shows the volume of the sound
- **Speechrate** is the rhythm of the speech, often presented as the number of syllables per second
- **Jitter and Shimmer** refer to frequency and amplitude perturbations respectively
- **Spectral centroid** indicates *the center of mass of the magnitude spectrum* [3]; it is associated with the brightness of a sound
- **Spectral flux** is a measure of the change speed of the power spectrum
- **Beat histogram** is a plot presenting the strength of different rhythmic periodicities in a signal
- **Beat sum** defines the regular beats in a signal
- **Strongest beat** is the strongest beat existed in a signal
- **Mel Frequency Cepstral Coefficients (MFCC)** are coefficients that form a mel-frequency spectrum (MFC). The term MFC is referred to a short-term power spectrum of a sound that approaches the human auditory system more approximately that any other linearly-spaced frequency band distribution.

The majority of the recent studies use a toolkit for audio feature extraction called OPENSMILE and it can extract all the aforementioned features [178]. Another commonly used software is the Praat [179].

Audio signal and emotions

Studies have already created some links between specific audio features and emotion. Increased levels of frequency combined with slower speechrate and high volume are associated with nervousness and agitation [34]. Moreover, high levels of volume with high frequency are related to fear [34]. Anxiety is characterized by higher pitch, voice tremor, several speech flaws, and faster articulation [180]. Sadness is associated with low pitch, long pauses, slow speechrate and soft voice whereas joy with high pitch, loud voice and faster speech rate. Interest is characterized by a large frequency range and fast-talking [34].

Body movements and gestures

Body movements usually complement other modalities, like speech or facial expressions, to facilitate social interaction. However, body language and gestures are also essential aspects of human communication, as they are human innate skills to directly decipher the different social signals.

Emotions and behaviors can be recognized by postures, whole-body movements, or gestures. For example, existing studies have shown that collar joint angle and shoulder joint angle can be used for such a purpose [181]. This has been also verified by another study where the amplitude of the elbow joint angles, combined with the head inclination, is associated with the expression of fear and anger [182]. The most commonly used features to evaluate the quality of movement are strength, fluidity, repetition, tempo, and amplitude [183].

In general, up to now, in HRI only simple actions have been examined [184]. Beck et al. examined some basic body postures using a Nao robot to see how humans can emotionally perceive them [185] . It is essential that robots can have integrated nonverbal aspects of communication as it can enhance their expressiveness.

Motion and emotions

Human body movements have been minutely meticulously described and interpreted by the method/language Laban Movement Analysis (LMA). It offers clear documentation of human motion and it is divided into four basic components: BODY which provides the structural and physical characteristics of the human body, EFFORT which describes the quality of the movement and the intention behind it, SHAPE which depicts the way the body changes during the movement and the SPACE, which connects the movement with its environment [186]. These movement characteristics are related to emotional states and studies are trying to discriminate human behaviors according to this emotion categorization. For example,

happy emotions are related to the spreading of movements whereas anger is represented by small, intense movements [35]. Fear is related to compressed and confined movements [35].

Physiological and Neurophysiological signals

Affective signals can represent the activity both of the peripheral and the central nervous system. For the peripheral one, the most commonly used signals are the ones coming from the electrodermal activity (EDA) and the cardiovascular activity such as heart rate (HR) [30]. EDA can be used as a direct measure of physiological arousal and it has been reported that there is a correlation with self-reported arousal [187]. The increase and decrease of HR, on the other hand, has been associated with several emotions, such as stress or happiness. There are more physiological signals used for the periphery, such as facial EMG signals or skin temperature.

Regarding the central nervous system, several neuroanatomical structures have been proved to be involved in the processing of affective information [30]. The most commonly used measure for this system is electroencephalography (EEG) due to its easiness of use and its low cost compared to other methods. EEG measures the electrical potentials from the brain and up to now, several characteristics have been associated with the emotional states. For example, the frontal alpha asymmetry, which measures the lateralization of brain function toward both frontal cortices separately, has been correlated with the valence dimension, meaning the range of positive to negative feelings. In general, the frontal lobe is more related to Valence emotions [188]. Moreover, an increase in fronto-central theta waves has also been related to detecting prosodic emotional changes [189] and thus, theta band is associated with the perception of emotions through vocal expressions [190].

Except from the affective states, delta oscillations play also an important role in cognitive processes like attention, memory, and decision making and they are focused on frontal, central, and parietal areas, as well as occipital if they are related to emotional processes [191]. Specifically, they have been associated with arousal in posterior brain areas and with valence in anterior brain areas, as well as with surprise [39]. Moreover, delta activity in frontal areas has also been linked to the perception of face recognition related to emotional expressions [192]. Physiological and neurophysiological signals have been used for the assessment of cognitive states, like the EDA we mentioned before which is associated with cognitive arousal and consequently with cognitive effort [30]. Another indicator has been considered to be the whole alpha activity, which presents a decrease towards an increasing arousal and an increase with relaxation [30]. However, studies are supporting that high-frequency bands, like beta or gamma, are more related to the Valence dimension compared to the lower ones [188, 193]. Unpleasant stimuli can trigger effects in the gamma range whereas other studies are supporting that higher frequencies are a reliable indicator of arousal

[39]. No matter what the band is, frontal and parietal areas have been proved to be the most dominant brain areas for emotion recognition [193]. The parietal area is also related to perception processes, and alpha activity in it acts as an index of presence experience while in a VE[194]. Posterior alpha activity has been associated with visual attention mechanisms [195] and thus it can be used as an indicator of the level of attention in a scene. Moreover, it has been noticed that prefrontal alpha and theta activity can vary according to different levels of cognitive effort and thus it can be used to measure mental engagement [30].

Psychophysiological studies have been started examining the interdependency of behaviors in human social interaction since the early 1980s. Researchers were trying to examine if the behavior of a person can affect or even predicts the behavior of others. Levenson and Gottman, for example, developed a method where they were measuring the coupling index from peripheral physiological signals and they found mainly negative social interactions [196]. Later on, the coupling index, also known as physiological linkage, was also correlated with empathy [197] showing that people who recognized the negative emotions of others, concluded to share their physiology. This was later explained by Janssen who supported that this is the outcome of emotional convergence, which is part of empathy and is mainly responsible for the adjustment of people's affective states through mimicry [168].

Two main systems have been identified to play an important role during a social interaction in humans: the mirror neuron system (MNS) and the mentalizing network (MTN) [30]. Recent neuroimaging studies have proved that social inferences from human interactions lead also to another brain network called the person perception network (PPN) [198]. The MNS was discovered by Giacomo Rizollati et al [199] and consists of structures like the premotor cortex, the primary somatosensory cortex, and the inferior parietal cortex. It is important for functions like action understanding, imitation, and empathy and its main role is to "mirror" the action of others. The second network, the MTN, represents our ability to decipher the mental states of others and it is important for the interpretation of other's intentions in novel or difficult situations [200]. It consists of the medial prefrontal cortex, the temporal lobes, the posterior superior temporal sulcus (PSTS), and the temporal-parietal junction.

EEG studies on individuals have already proved that MNS plays a role in social interactions. However, recent studies have started to examine the interaction between humans and robots or avatars to verify if and what is the difference in these systems between H-H and H-NH interactions. Urgen et al. [201] examined the difference in perception between humans and robots (mechanical and anthropomorphic) having the participants watch some video clips of actions performed by three agents: a human, a humanoid and a mechanical robot. They used only EEG recording for examining the sensorimotor mu rhythm (8-13 Hz) which is linked to the motor simulation aspect of action processing and acts as a MNS index, as well as the frontal theta (4-8 Hz) related to semantic and memory-related aspects. Their results showed that the human

MNS cannot differentiate the actions of robots and humans and thus its activity is similar. Moreover, frontal theta oscillations were noticed when humans interacted with the humanoid, but not with the mechanical robot. Yoon et al. examined human-robot interaction under the context of meditation, and they noticed fewer gamma and beta activity, especially in the frontal area, for the group that was practically interacting with the robot compared to the one that was just listening to it [202]. This reveals that the people who were guided by the robot were in a more relaxed state.

Wang et al. [198] also examined such an interaction based on the observation of images of several social interactions, presented in two versions: human-human and human-robot. They used a Nao mechanical robot and fMRI to detect evoked emotions and differences in neural processing between the two states. The outcome was that robot observation leads to lower MTN engagement and thus, the interaction between the robot and the human was considered less believable.

2.6 Modeling, Analysis, and Synthesis of Human Behavior

Data can be collected by scripted or non-scripted scenarios. The majority of the research is conducted based on predefined scenarios that however, lack naturalism and even spontaneity in users' responses. Some studies have tried to collect data from real-world situations, like real phone conversations [27] but they consist of enough limitations as they are based on unimodality and the quality of the recordings still remain low. Thus, the collection of data able to be used in multimodal research requires a combination of a real-world scenario, naturalistic setup and resources, as well as equipment to ensure the high quality of the recordings. This leaves the option of a laboratory setting where the setup is carefully designed and controlled. Datasets of recent studies include physiological signals, motion capture, and computer vision and act as an effort to create models of realistic human social behavior [203–206]. Table 2.2 summarizes the most commonly used multimodal datasets created from dyadic human and nonhuman interactions or other affective stimuli, like music.

Table 2.1 The most commonly used multimodal datasets up to now on realistic human social behavior.

Reference	DATASET	# of part.	Sensors	Modalities	Type of interaction	Result
Cafaro et al. *[203]*	NoXi	84	Kinect Headset	Gestures, Facial behavior,	Video HHI	A multi-lingual database of natural dyadic expert-novice interactions, focused on unexpected events

				Audio (prosodic and acoustic features)		
Bilakhia et al. *[167]*	MAHNOB-Mimicry	60	Headset+far field microphone, Cameras	Audio, Face and head movements	Natural HHI	A set of highly-accurately synchronized multi-sensory audiovisual recordings of naturalistic dyadic interactions, for the study of mimicry and negotiation behavior
McKeown et al. *[204]*	SEMAINE	150	AVT stringay cameras, Head and room microphone	Significant gestures (head shakes and nods) and facial actions, speech and prosodic features	Video HCI	A large audiovisual dataset derived by emotional enhanced conversation between a human and a Sensitive Artificial Listener (SAL) agent
Koelstra et al. [206]	DEAP	32	Biosemi ActiveTwo system, Peripheral physiological sensors, Sony DCR-HC27E camera	EEG, facial EMG, EOG, GSR, temperature, respiration, self-assessment questionnaire	Video, HCI	A database of spontaneous emotions induced by music
Douglas-Cowie et al. [205]	HUMAINE	125	Camera, Microphone, Physiological sensors,	Speech and language, Gestures, Facial actions, Physiological measures (ECG, breathing, GSR), questionnaire	Natural and Video HHI HCI	A dataset of naturalistic and induced emotions in several contexts
Ringeval et al. [207]	RECOLA	46	Video, Microphone, Physiological sensors	Audio, Body language, ECG, EDA, Self-assessment questionnaire	Video HHI	Dyadic collaborative and Affective Human Interactions
Lefter et al. [208]	NAA	16	Microphone, Microsoft Kinect v2, Peripheral physiological measures	Audio, Gestures and body movements, EMG, ECG	Natural HHI	A dataset of dyadic interactions for negative affect and aggresion

Newman et al. [209]	Harmonic	24	Video, Physiological signals	Eye gaze, Arm joint positions, EMG,	Natural HRI	A dataset of human interactions with a robotic arm measuring mental states and intention
Hazer-Rau et al. [210]	uulmMAC	60	Microphone, Kinect v2, Physiological signals	Audio, Body movements, EMG, ECG, SCL, respiration, body temperature, questionnaire	Natural HCI	A dataset for emotional and cognitive states recognition in a mobile interactive HCI gaming scenario

Several different studies have tried to examine human behavior during several kinds of interactions, aiming to model the human social behavior in order to enhance the domain of H-NH interaction. An interesting study is the one of Rasheed et al. [211] who examined dyadic face to face dialogs, extracting audio features to assess speaking mannerisms and human social behavior that can be used as a real-time sociofeedback system. The researchers extracted nonverbal speech cues, including conversational and prosodic features using HMM and based on their manual annotation they concluded to some speech mannerisms. Then, they created the link between the aforementioned features and human social behavior. Another study where the examination of nonverbal human behavior was used for an expressive virtual tutor is the one of Ben Moussa et al [212] . They extracted audio signals with nonverbal cues and facial expressions and they modeled the nonverbal behavior in order to create a complete system of virtual tutoring.

There are also enough studies that have tried to decipher human behavior completely from a clear human perspective, without directly aiming to apply their results to the technology. From this point and on, the question raised is how we can do translate all these features to the broad HRI field or how we can create a better, more efficient connection between humans and agents. As we are in the golden age of humanizing social agents, there are a lot of thoughts and concerns over this topic trying to find answers on how, and most importantly if, humanlike agents can improve HCI and HRI and facilitate human acceptance [213]. Figure 2.4 is a table from a recent study of Giger et al. [213] where the authors have collected all the negative and positive aspects of robotization and the effort of making social agents resemble human behavior.

Humanization type	Positive aspects	Negative aspects
Psychological	• Interaction engagement • Wellbeing benefits • Educational benefits • Increased motivation • Higher perceived support • Increased social connection	• Overtrust and unrealistic perceptions of a robots' autonomy and capabilities • Attachment issues • Existential threat
Physical	• Increased social interaction • Higher perceived assistance • Higher proximity	• Feelings of eeriness or discomfort
Functional	• Economic gains • Frees humans from dull tasks • Frees humans from dangerous tasks • Increased precision (e.g., health), and reaching places otherwise inaccessible (e.g., deep sea; space, disaster exploration)	• Unemployment • Requires human supervision • Creates demands for the acquisition of new skills (e.g., doctors who work with surgical robots need to know how to operate the robots).

Figure 2.4 *The most recent table, created by Giger et al. [213], expressing the thoughts and concerns of humanizing social robots*

2.6.1 Robotic Psychology

Stock et al., in their very recent work, presented "robotic psychology", which aims to find and cover the gap between humans and robots by shading some light in features peculiar to HRI and, more broadly, to HCI [21]. Statistical researches expect that the number of social and service robots will increase dramatically, integrating around 1.2 million robots in domains like medicine, public relations, etc., and thus, it is important to orient the research towards humans extracting human values and applying them to the technology. The latter can ensure a more successful and efficient collaboration between humans and robots or other kinds of similar technology (i.e avatars) [21].

Robotic psychology, enhancing the field of H-NH interaction and going a step further, tries to examine the relation of humans and robots via a sensorimotor, emotional, cognitive, and social level [214]. The understanding of human responses in all the aforementioned domains can help us decipher the human needs towards technology and to verify where we stand up to now. Stock et al. modeled the framework of this concept, as shown in Figure 2.5, to facilitate the description and exploration of humans' behavior in the context of a technological environment [21].

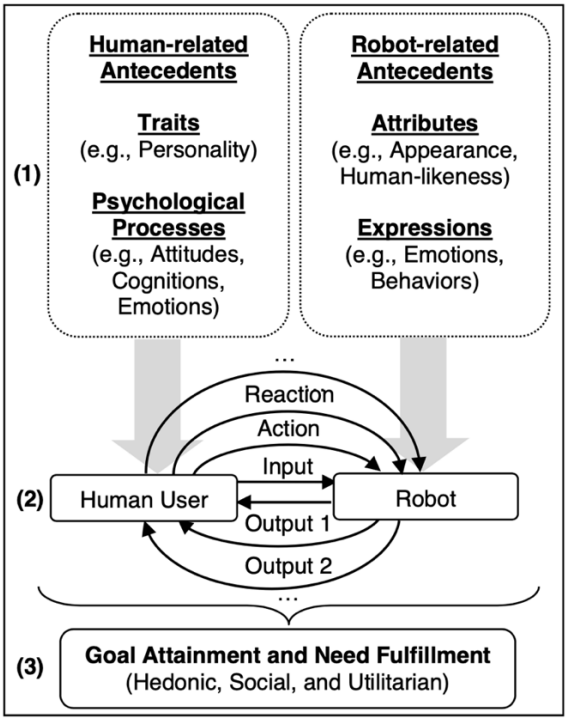


Figure 2.5 The conceptualization of robotic psychology as modeled by Stock et al. [21]

The purpose of this model is to reveal human behavioral patterns that need to be clearly differentiated by experiences during HCI or HHI. We can see from Figure 2.5 that robotic psychology consists of three levels: the Individual, the Interaction, and the Outcome level. At the individual level, personal features are examined, either human-related or robot-related, to extract the behavioral patterns. During the Interaction level, features and effects that are extracted through the mutual influencing processing are examined. Normally, the robot should receive the human reaction as input and respond in a corresponding manner. Towards this direction, nowadays, several social robots are being developed to detect humans' affection and react to it [215–217]. The Outcome level mainly controls if the interaction was successful by fulfilling the predefined needs.

To facilitate such kind of research, studies have examined several factors regarding how humans evaluate technology and how the latter affects humans. Humans, with their multidisciplinary mature, have a lot of factors that need to be considered for a successful H-NH Interaction. Personality is one of them, but we need also to assess attitudes and provoked emotions. By attitude, we refer to a mental and neural state of promptitude which is modified through experience and for its assessment there are several verified questionnaires, like the Negative Attitudes Toward Robots Scale (NARS) that Nomura et al. designed [218]. In general, such scales are designed to assess three levels of attitude toward interaction with robots, social influence, and emotional experience with them. Up to now, negative attitudes have been associated with

specific human behaviors like emotional expression constraint, avoidance of touching, and lack of communication [32, 218, 219].

Emotions, on the other hand, play a major role in the engagement during H-NH interactions and they can act as predictors of human behavior. In the context of HRI, the most common emotion detected is anxiety and it is correlated with avoidance or distancing between humans and robots [21]. This emotion was also verified by the work of Lupkowski and Gierszewska, who tested the UV effect in humanoid characters, and they found that when the UV effect was apparent, the dominant emotions were anxiety and strangeness, whereas the highest comfort level was noticed for a cartoon-based character [32]. It is said though that emotions are shaped based on previous experiences a human may have with other humans or other kinds of technology [21]. However, from another point of view, it is very important to examine how robots can "transfer" emotions to humans, gaining better access to them. Stock et al. used a term for this, as "*emotional contagion*", referring to any way robots can use, mainly mimicry and synchronization of humans' features, to express but also transfer emotions [21]. Emotional processes, as shown in figure 2.6, can be conscious or subconscious and can serve or fulfill different purposes.

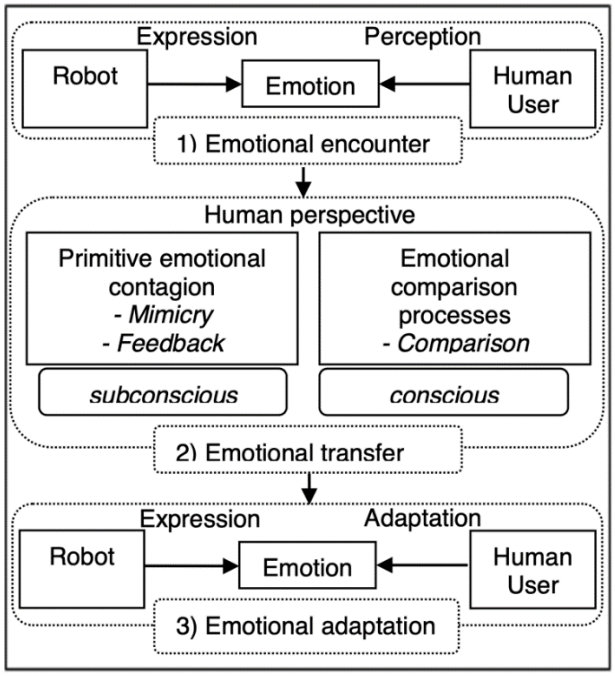


Figure 2.6 Diagram depicted the emotional contagion in HRI, as designed by Stock et al. [21]

Table 2.1 Studies on human perception and behavior under several types of interaction during the last decade

RR: Robot with realistic appearance, HAI: Human-Avatar Interaction, PP: Physical presence

REFERENCES	PURPOSE	FEATURES										RESULTS
		HRI	RR	HAI	HHI	PP	EEG	EMG	Voice	Body	Questionnaire	
Mollahosseini et al. (2018) *[62]*	The role of embodiment and presence in human perception of agent's facial cues	✓	×	✓	✓	×	×	×	✓	×	✓	The eye gaze and some facial expressions are perceived better when the embodied agent is physically present
Mara et al. (2020) *[81]*	Correlation between agents' voice and appearance's expectation	✓	✓	×	×	×	×	×	✓	×	✓	The more human-like the voice, higher the expectation of anthropomorphism
Li et al. (2016) *[16]*	Comparison between a VH and a social robot as a video instructor in an educational context	✓	×	✓	✓	×	×	×	×	×	✓	The preference is towards the human lecturer, however agents, if designed well, can provide an alternative with a higher preference in the robot.
Urgen et al. (2013) *[201]*	Brain theta and mu activity during human-robot interaction	✓	✓	×	✓	×	✓	×	×	×	×	A robot with mechanical appearance results in a greater frontal theta activity which is correlated with greater memory processing
Hofree et al. (2014) *[45]*	Differences between physically present and virtual android in humans' mimicry	✓	✓	×	×	✓	×	✓	×	×	✓	The physical interaction made the users feel more uncomfortable. However, mimicry occurs naturally compared to the virtual presence.

Birmingham et al. (2020) *[42]*	Use of a Nao robot as a mediator in a support group for control stress	✓	×	×	✓	✓	×	×	×	×	✓	The robot made the discussion mechanical, with lack of real flow (lack of faxial expressions and non natural voice)
Inoue et al. (2021) *[13]*	Comparison between an android (ERICA) and a virtual agent for a job interview training	✓	✓	✓	×	×	×	×	×	×	✓	Similar results for both agents, but the virtual one lacks the physical presence
Lupkowski et al. (2019) *[32]*	Evaluation of the UVH and of the emotional response to the humanoid models	✓	×	✓	×	×	×	×	×	×	✓	Individuals' attitudes can influence their perception towards agents. The higher the belief in human uniqueness, the higher the nervousness towards artificial agents
Rasheed et al. (2013) *[211]*	Use of conversational and prosodic features to assess human social behavior	×	×	×	✓	✓	×	×	✓	×	✓	A real time system that combined sociometrics with speech features representing human behavior in dialogs
Moreau et al. (2019) *[220]*	Brain correlates during Human-Avatar joint performance	×	×	✓	×	×	✓	×	×	✓	×	Fronto-central and occipito-temporal theta activity for processing and integrating visual and motor information in social interaction
Marschner et al. (2015) *[73]*	The role of body and gaze direction on attention and emotional responding in social human-avatar communication	×	×	✓	×	×	×	✓	×	✓	✓	The gaze and the body direction of a VA can ifluence the visual attention of the human, the facial mimicry and the experience

Shiban et al. (2015) *[54]*	How the appearance of a virtual agent can influence the performance and motivation in an education context	×	×	✓	×	×	×	×	×	×	✓	Appearance features can influence independently the performance and motivation of students→proposition for a personalized tutor. The younger and more attractive avatar had a more positive impact.
Wauck et al. (2018) *[157]*	How the appearance of a virtual avatar can affect the performance in a game context	×	×	✓	×	×	×	×	×	×	✓	Self-relevance cannot influence the performance of the user in a gaming context and has no effect on gender
Yokotani et al. (2018) *[75]*	Comparison between clinical psychologists and virtual agents for a mental health interview	×	×	✓	✓	✓	×	×	×	×	✓	The anonimity of the VAs is relevant to patients' self-disclosure
Amershi et al. (2019) *[221]*	Verification of design guidelines for human-AI interaction	×	×	×	×	×	×	×	×	×	✓	Proposition and evalutation of 18 desing guidelines for human-AI interaction
Our work	A multimodal in-depth documentation, analysis, and comparison between H-H and H-NH interactions	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	Differences in human perception, behavior and emotions, a new database with human reactions extracted from H-H and H-NH interactions, a voice-based model differentiating H-H and H-NH interactions

Table 2.3 above presents a selection of studies that have examined human features and human behavior via several modalities under different types of interaction during the last decade. The modalities and the interactions described in this table were selected based on the setup of our work. The first part of the table presents studies that have used at least HRI, the second part HHI, and the last part HAI. Summing up, we can see that robots with human-like appearance and physically present, are more difficult to be perceived and they don't easily allow humans to feel comfortable. However, mechanical features like lack of facial expressions or non-natural voice create a distance. Thus, it is important to find if there is a balance between the human-likeness and humans needs. Up to now, most of the studies have shown that the most important features a robot or a digital human should have to be more acceptable by humans are the facial expressions and the eye contact. Moreover, in studies where robots and virtual humans have been compared, the former has been preferred, mainly because of its physical presence. Some studies have already tried to extract humans' features only during H-H social interactions. However, to integrate this information into nonhuman agents, we need a direct comparison between H-H and H-NH interactions to evaluate what is really missing. In such a case, the more features we extract (physiological, psychological, technical), the better we will understand human behavior, and thus, the better we can approach the design and the functionality of the nonhuman agents. This direct multimodal comparison is one of the basic limitations of the current literature.

2.7 Summary and Discussion

In this chapter, we tried to cover all the dimensions of H-NH interaction and the ways of approaching them. We analyzed the role of human likeness and presence in human perception, concluding that human expectations are directly linked to the influence a robot's appearance may have. A human-like voice, for example, should be accompanied by an anthropomorphic appearance. We described the research behind human perception during HRI and the role of social robots and virtual humans in human life. We then presented in detail any physiological and psychological modality used to capture and measure human behavior during social interactions, concluding that EEG is one of the most reliable measures.

Inspired by all described approaches, our work aims to fulfill current limitations, as described in chapter 1, and provide some answers regarding human perception and humans' behavioral and attitude patterns towards social nonhuman agents developed based on the up-to-date technology. Our goal is to delve deeper into the features and the nature of H-H and H-NH social interactions, providing a detailed validated assessment of how humans react towards technology but also how the latter affects humans. We are trying to decipher the complex human social behavior by measuring as many features as possible, like audio,

motion, upper body muscles, brain signals as well as psychological indexes. Psychometric measures can help us verify the provoked emotions, personality traits, or even personal attitudes. Our work is based on a human-human, human-avatar, and human-robot natural interaction under the same scenarios that allow us to proceed to a direct comparison between the three different conditions, which is clearly missing from the up-to-date literature. People are more used to the existence of animated on-screen avatars, but they are unfamiliar with the presence of physical robots and this can provoke biased responses. To examine this unfamiliarity we conducted another experiment where we studied H-H and HR interactions, using a human and an identical humanoid robot with which participants had to interact under the same scenario.

Our main question after all is to what extent a social agent or a robot need to be human-like to fulfill humans' needs and to be socially accepted. We need to find the key point where humans socially approve and accept social agents and social agents have as a principle the human needs. Several studies are trying to verify the UVH and the role of human-likeness in human-nonhuman interaction but as Katsyri et al. [55] and Ratajczyk et al. [219] noticed, there is a lack of physiological human measures that can verify all the used well-validated questionnaires dedicated to this purpose. We tried to cover this limitation by using a multimodal approach. Moreover, to complement the above, we tested our humanoid robot, Nadine, under several roles in order to examine when and how the human-likeness can affect our perception, imagination, emotions, or even concentration and how different roles can affect human's preference.

We want eventually to create a better link between humans and technology that can facilitate the use of social agents in several fields like health or education. Our purpose is to use human's complete behavior to reveal the humans' needs towards human-nonhuman communication and to find a possible way to suggest potential improvements. Our approach might also define what is still missing from the existing technology, trying to cover some of the negative aspects of Figure 2.4.

CHAPTER 3

INTERACTION IN VIRTUAL AND PHYSICAL ENVIRONMENTS

"Virtual Reality is a self-created form of chosen reality. Therefore it exists."

-Joan Lowery Nixon, American journalist and author

Interaction in Virtual and Physical Environments

3.1 Introduction

VEs, due to their multi-sensory nature, have become a very powerful tool in several domains and nowadays, people tend to prefer their use over a simple 3D environment. The question though is if there is indeed a need of using VR and whether the exposure to VR applications can affect the brain activity of the users and their cognitive, behavioral, motor, or other functions, or VR has just started to be considered "in fashion". However, we don't underestimate the usefulness of VR as it is undoubtable that it can expand the possibilities of the real world. Moreover, in the context of an experiment setup it can ensure that the conditions and the parameters used are controlled and adjustable.

To address the limitations of previous studies, we conducted a small experiment with three different types of environments. It is said that VR engages the sensorimotor system, so we wanted to verify if it can provoke naturalistic psychological and behavioral responses. It has been proved that VR can simulate or rehash the neural activity induced by a Physical Environment (PE) and thus, VR platforms can be designed in such a way whereby they can create the desired and adequate differences between neural functions in PE and VE [222]. Therefore, through VR the quantification of the parameters becomes easier.

The thought behind this study was to examine if there is indeed a need of using VR and what is the optimal environment for a digital human. We need to mention that VR depends on the up-to-date technology so it is sure that a lot will change when technology becomes more advanced. Our study is conducted with the up-to-date technology and aims to answer questions of nowadays.

Thus, in this chapter, we present the experimental design, the methodology, the results, and the total outcome of our first experiment.

3.2 Experimental Design

We conducted an experiment with three different types of environments: Physical/Real environment (PE), VR with realistic design (VR) and VR with imaginary design (VRI), as illustrated in Figure 3.1. The reason we chose two different VR environments was to examine if the graphics and the design of the environment can play a role in human perception and maximize the impact. The scenario for the three cases was the same and it is based on a classroom environment where multiculturalism and bullying cases were taking place. During the experiment, the users could experience the environment from a teacher perspective but

also through the eyes of the student who received the bullying. We expected that the observation of the scene from both perspectives would provoke a higher sense of engagement and empathy. No physical interaction was expected from the participants as the main goal was to evaluate the impact of the environment. The total duration of the experiment was about 2-3 minutes for each case.

Volunteers were separated into three groups, according to the different environments. The VR group depicted a realistic classroom, similar to the one we used for the physical environment. The VRI group was exposed to an imaginary class environment, different from the setup that people with academic background are used to. The third group of the PE was used to validate the effect of the VR. The whole procedure took place at the University of Geneva (UNIGE). The scenario and the dialogs for all three cases were identical.

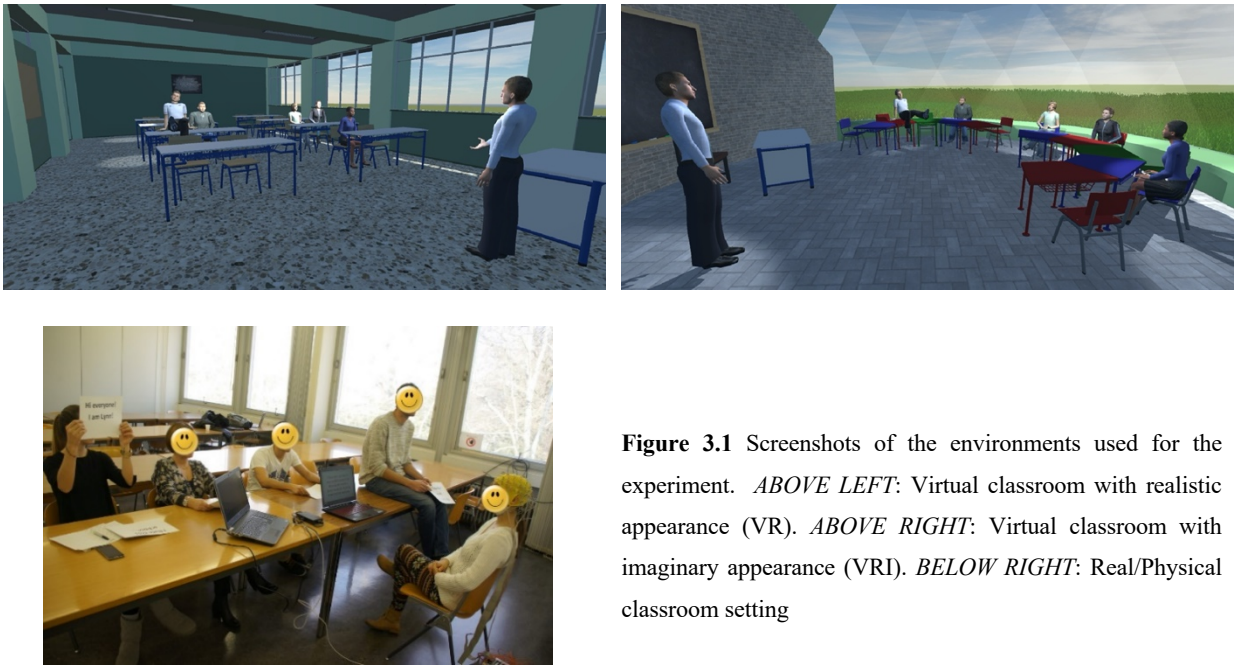


Figure 3.1 Screenshots of the environments used for the experiment. *ABOVE LEFT*: Virtual classroom with realistic appearance (VR). *ABOVE RIGHT*: Virtual classroom with imaginary appearance (VRI). *BELOW RIGHT*: Real/Physical classroom setting

The environments were developed using the Unity3D game engine in collaboration with the Visual Computing Media Lab at Cyprus University of Technology (CUT). The avatars used were created with the Maya Autodesk Character Generator and the methodology is detailed in [223]. To ensure the sense of presence participants had the experience of the Head Mounted Display VIVE.

For the validation of our experiment and the proper examination of the VR effect, we used EEG recordings to capture the brain activity, as well as psychometric measures through a questionnaire that subjectively examined the sense of presence, the empathy, and the reflections of the users.

3.2.1 Participants

Thirty-three healthy adults (22 males and 11 females), aged from 25 to 59 years old participated voluntarily in this study. All our subjects were Ph.D. students, Postdoctoral researchers and Professors currently working at the UNIGE, as the academic profile was necessary. We ensured that the majority of our participants had no previous experience with VR (only 6% claimed an experience with such environments). All participants were fully informed about the whole procedure and the time-consuming placement of the EEG device, for which we made sure that no discomfort was presented.

3.3 Data Collection and Analysis

3.3.1 EEG recordings and Analysis

EEG signals were recorded and amplified using a BIOSEMI Active Two 64 channel amplifier system (www.biosemi.com). Active electrodes were used in association with a headcap, on which 64 electrodes were attached according to 10-20 system at the locations Fp (Fp1, Fpz, Fp2), AF (AF7, AF3, AFz, AF4, AF8), F (F7,F5, F3, F1, Fz, F2, F4, F6,F8), FT(FT7, FT8), FC (FC5, FC3, FC1, FCz, FC2, FC4, FC6), T(T7, T8), C (C5, C3, C1, Cz, C2, C4, C6), TP (TP7, TP8), CP (CP5, CP3, CP1, CPz, CP2, CP4, CP6), P(P9, P7, P5, P3, P1, Pz, P2, P4, P6, P8, P10), PO(PO7, PO3, POz, PO4, PO8), O (O1, Oz, O2), I (Iz) based on BIOSEMI layout and referenced to Cz. BIOSEMI's acquisition program, ActiveView, was used to record the data with electrode impedance < 5kΩ and sample rate 2048 Hz.

The analysis of the EEG data and the processing of the signal were carried out in MATLAB. All data were carefully checked for artifacts, like eye blinks or abrupt head and/or body movements. Raw signals were filtered using a bandpass filter from 0.1 to 60 Hz. The electrical line noise (60Hz) was removed using a notch filter. Signal was segmented over time windows of 20 seconds and Fast Fourier Transforms (FFT) was applied to each segment for each electrode so that we could transfer and analyze it to the frequency domain. Then the power spectra were calculated, as well as the average across segments to examine the total activation in the ten selected Regions of Interest (ROIs), as depicted in Figure 3.2, enabling us to reveal the dominant frequency for each brain area. We also chose to examine the occipital region (10^{th} ROI), as it is mainly associated with the visual attention mechanisms. This can give us some results concerning the difference between the designs of the two VE considered in the experiment. Brain rhythms, which we are mainly interested in, are theta (3-7 Hz), alpha (8-12 Hz), beta state (13-30 Hz), and low gamma (30-42 Hz). The time that the brain needs to adapt to the new state, was also calculated based on the activity seen in the segments.

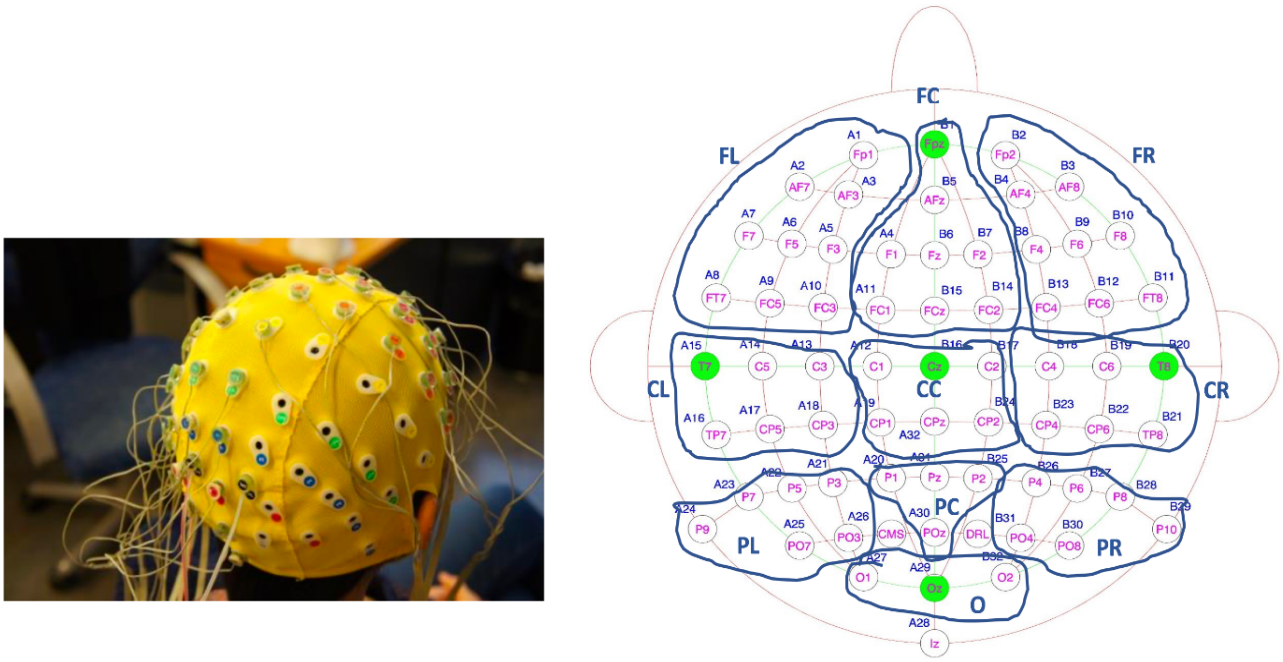


Figure 3.2 LEFT: A participant weating the headcap with some EEG electrodes on. RIGHT: Regions of Interest, separated according to the brain areas we wanted to examine: FL = Frontal left, FC = Frontal center, FR = Frontal right, CL = Central left, CC = Central center, CR = Central right, PL = Parietal left, PC = Parietal center, PR = Parietal right, O = Occipital

The participants were informed on how to minimize the inducing noise into the ongoing EEG signal, e.g they were advised not to make big movements in order to avoid wrong comments on the signal. Moreover, for this reason, the experiment was also conducted in a sitting position, so that the subjects didn't have any distraction from the cables or other EEG equipment. The environment was also carefully modified in order not to influence the subject and measurement system, e.g no fluorescent lamps were near the system.

Brain signaling and Brain waves

Several neurophysiological methods have been used to monitor and assess brain activity, the choice of which depends on the needs and the design of each research. Electroencephalography (EEG), thanks to its design, constitutes the most comfortable, but most importantly the most cost-effective non-invasive solution.

EEG measures the cortical activity and specifically it measures the summed electrical field potentials from cortical neurons that correspond at a specific region of the scalp to each electrode used [**Error! Reference source not found.**]. It has been proved that certain types of neurons have oscillatory properties that can reveal rhythmic EEG activities [**Error! Reference source not found.**]. Thus, every time a neuron in the brain is activated during synaptic excitation of the dendrites, creates a local current that can be measured

as an EEG signal. The dynamic pattern of synchronization and desynchronization within different neuronal groups generates the visible outcome of the EEG. Any change in the level of synchronization of such activity is depicted in changes in EEG amplitude of several frequencies at the scalp [224].

Frequency band	Frequency	Brain states
Gamma (γ)	>35 Hz	Concentration
Beta (β)	12–35 Hz	Anxiety dominant, active, external attention, relaxed
Alpha (α)	8–12 Hz	Very relaxed, passive attention
Theta (θ)	4–8 Hz	Deeply relaxed, inward focused
Delta (δ)	0.5–4 Hz	Sleep

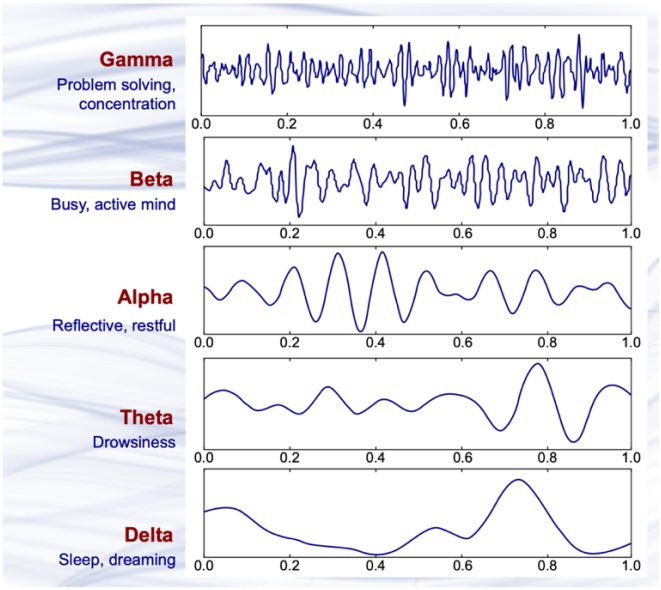


Figure 3.3 ABOVE: The frequencies and the characteristics of the five basic brain waves as we use them today, described by Abhang et al. [225]. BELOW: Samples of the five basic brain waves in the time domain [225]

These rhythms have been translated through the so-called brain waves and can correspond to a brain state. Nowadays, we have concluded to the five basic brain waves, whose main frequencies, for a human EEG, are shown in Figure 3.3, as described by Abhang et al. [225]. The alpha rhythm was noticed and named first, in 1934 by Andrian and Matthews [226].

Each brain area is related to a different brain function, and consequently body function. Usually, according to the dominance of the brain waves in each region, we can figure out how the latter is activated and what is its role in the examined task. However, the exact location of each region is not always found and this is why EEG has still some limitations as a technique [225]. Another limitation is its low spatial resolution and

its lengthy set up procedure [225]. Albeit all these, it allows the examination of fast temporal dynamics and it remains the most accessible neurophysiological method for modern research.

3.3.2 Psychometric data

To complete our physiological measures, we used a validated questionnaire named Igroup Presence Questionnaire (IPQ) [227], including 14 items for the measurement of presence and the perception of the environment.

3.3.3 Statistics

Reliability tests were carried out for the variables for both EEG and The IPQ. The overall alpha was 0.715 > 0.7 for the questionnaire and 0.893 > 0.7, as well as 0.719 > 7 for the brain frequency and the time accordingly. This indicates the reliability of our variables.

However, the tests of normality used (Kolmogorov-Smirnov Test and Shapiro-Wilk Test) indicated that our data didn't follow a normal distribution. Thus, we used non-parametric tests (Mann Whitney and Kruskal Wallis tests) to validate the significance level.

3.4 Results: Evaluation of Human Perception between Virtual and Physical Environments

Our aim in this study was to examine the possible influence of a VE in human brain, helping us understand how humans perceive the technology of VR. To wit, we were interested to see if the exposure in a VE can affect several functions of a healthy brain compared to an exposure to a physical real environment.

Minutely, we examined the dominant frequencies in order to find the dominant brain state, in different brain areas, examining the possible influence that a VE might have over a PE. Moreover, we compared the two different VEs investigating if the features of each environment can play a significant role in influencing the brain of the users. Finally, we measured the average time needed for the adaptation of the brain to the new dominant state, defining the optimal duration of a task in order to be effective. To validate our results, apart from analyzing EEG signals, we also used an Igroup Presence questionnaire to examine if the way we think that we perceive the environments is in line with the reaction of our brain.

3.4.1 Effects of VR in specific regions of the brain, influencing behavioral, motor, or other functions

First, we examined each selected brain area separately. In Frontal area, we found that both VR groups were synchronized in a theta state. Specifically, VR group reached an average frequency of around 6 Hz, for both hemispheres. VRI group was also synchronized in a theta state, but in a slightly higher value. Frontal theta state has been associated with focused attentional processing and especially cognitive effort and novelty detection [198]. In general, the frequency in the theta power band increases as the relevant task is becoming more demanding [201]. On the other hand, for the PE group we noticed a low beta state, which is linked with memory recall for this part of the brain. In this group, participants had just to deal with conditions they face in their everyday life. Thus, the beta state can be completely explained, as the procedure probably triggered images extracted from their memory, without the need for an extra cognitive effort. To the best of our knowledge, no theta activation has so far been observed in frontal regions during a VR experience.

Regarding the central area, it was difficult to derive decisive conclusions because the associated data was noisy. Moreover, as we will mention later, we noticed that the signal on this part of the brain needed the longest time to adapt in a state. However, we found that both VR groups were mainly synchronized in alpha state. The presence of such a state in this brain area has been linked to creativity-related demands and the process of producing new ideas [193]. In our case, this can be fully explained by the context of our educational scenario. Nevertheless, as we can see in Figure 4, VRI group showed a clearer alpha dominance, which means that more creativity-related mechanisms were engaged. However, for PE we found a clear beta state dominance and the difference between PE and the two VR groups presents a high statistical significance.

As far as the parietal area was concerned, we found exactly what we expected. Alpha rhythm in the parietal area has been used in several studies as an indicator of the sense of presence in a VE [194][228]. Thus, we confirmed the dominance of alpha power, compared to the PE group for which we noticed a beta state. The sense of presence was also verified by the results of our questionnaire where both VE groups reported a high level of presence. Moreover, the alpha band in this area is mostly connected to perception processes [201].

Lastly, we examined the activation of the occipital area that is associated with visual attention mechanisms and the recognition of objects. Interestingly, only the VRI group showed a synchronization in the alpha state. Although the difference between VR and VRI is not significant, we could explain such a difference due to the familiarity the simulated physical environment may have. In the VR group, participants didn't

have to process new features in the overall environment given that it was almost identical to the physical one. The difference between VRI and PE though was highly significant.

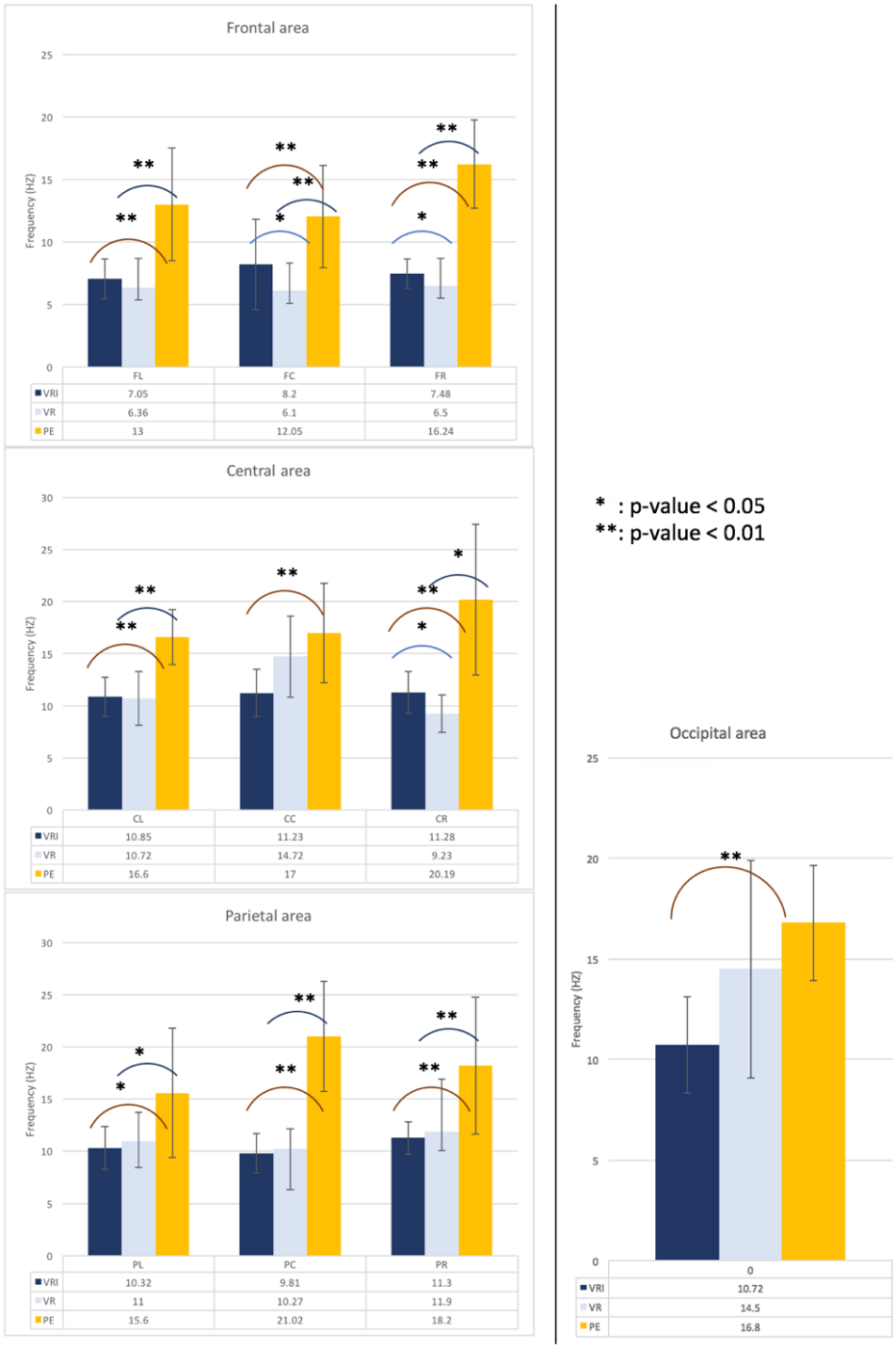


Figure 3.4 Power spectra observed in each of the 10 ROIs, in response to each environment. The 10 ROIs have been depicted in Fig. 3.2. Power spectra was defined as $PS = abs(filtered\ signal)\text{^}2$. Frequency bands examined: theta (4 – 7.9 Hz), alpha (8 –

12.9 Hz), beta low (13 – 20) and beta high (20 – 30 Hz). We didn't notice any activation in the low gamma band (30 – 45 Hz). Whiskers indicate the standard deviation and the asterisks indicate the level of significance: *p < 0.05, **p < 0.01 according to Mann Whitney and Kruskal Wallis test.

In general, the increase in alpha power in posterior brain regions can be an indicator of qualitative information processing [193], so the fact that we noticed it in both parietal and occipital lobes for both VR groups demonstrates a possible efficient recruitment of the desired networks.

Figure 3.4 shows explicitly the average frequencies for all brain areas and both hemispheres, indicating the dominant brain state for each and the significant differences. We can notice that we always have a significant difference between the groups exposed to the VR settings compared to the one having had the experience of the physical classroom. Based also on our questionnaire results, we found that both VR groups reported high levels of presence, although the one with the realistic design claimed a higher level. Moreover, we were able to create a link between the spatial presence and the experience of the user. Spearman's correlation provides us with a positive correlation between this feature and the participants' previous experience in VR, indicating that the more experienced users are in VR, the higher level of presence they can meet.

Summing up, we found that VR, by itself, without any obvious interaction of the participants with the environment, can indeed affect functions of the brain, showing different effects in different brain areas. This means that it can be used for several behavioral, motor, cognitive or other needs.

3.4.2 The role of the VR design

Although we did not find many significant differences between the two VR groups, they have been enough for a first conclusion regarding the role of the design of a VE.

The biggest difference between the two groups is located in the occipital lobe and can clearly indicate the distinction between the two VEs in terms of visual attention. Moreover, the slight difference that the two groups presented in the central area, with the VRI group showing a higher alpha dominance, can also reveal the engagement of more creativity-related mechanisms.

In general, the main differences we found between the two VE groups were in frontal and occipital lobes, indicating that the design and the graphics of the environment do play a role in the brain effect. We also noticed that the VR group needed slightly less time to process the information and adapt to the new state. However, only the frontal region of both hemispheres, the left part of the parietal lobe and the occipital lobe showed a significant difference. In any case, this can be attributed to the familiarity provoked by the VR setting compared to the one with the imaginary design.

We extracted also some interesting results through our questionnaire. The participants of the VR group declared that they were more aware of the real environment around them and any possible external stimulus, compared to VRI. This can also be directly correlated with the impact of the design of the VE.

3.4.3 Brain adaptation time to a VE

Figure 3.5 shows the time needed for the adaptation of the brain to the new state for each brain part for both VE groups separately. We noticed that, in general, the VR group needed less time to process the given information and proceed to the adaptation.

Given that we don't have a lot of significant differences between the time needed for each brain area, except the frontal region in both hemispheres, the left part of the parietal lobe and the occipital lobe, we calculated the average of the time between both groups, concluding that the mean time the brain needs to perceive a VE is 42.8 seconds (SD = 4.1 seconds). This means that a VR task should ideally last for at least 43 seconds so that the brain understands the impact of the VR and so the task can be effective.

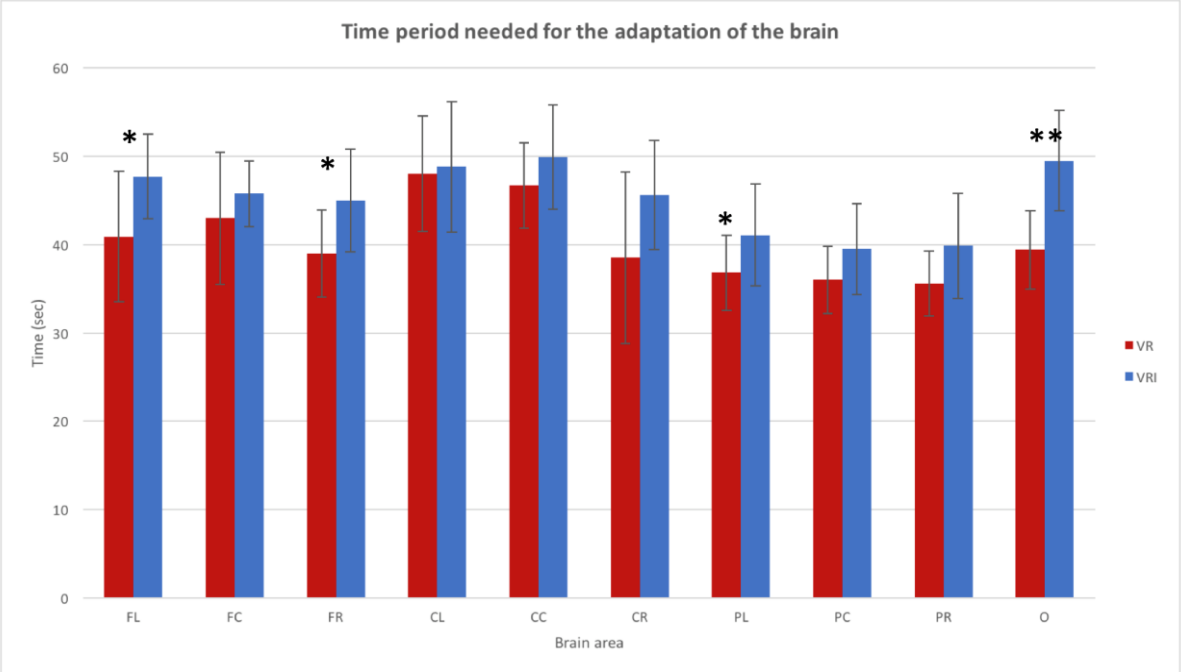


Figure 3.5 The duration of time needed for the adaptation of the brain to the new state for both VR groups. Significant differences can be observed in frontal and occipital regions. Whiskers indicate the standard deviation and the asterisks indicate the level of significance: *p < 0.05, **p < 0.01 according to Mann Whitney and Kruskal Wallis tests.

3.5 Summary and Discussion

With this study, we explored if exposure to VEs can affect the brain and how, compared to exposure in physical environments. We also used two VEs, a realistic and an imaginary one, to test the effects of different VR features. To verify our results, we formed a third, control, group which followed the same task in a real environment. Brain signals and a questionnaire were used for validation.

With this study, we answer our first research question: *Can a simulated experience, like Virtual Reality, activate regions of the brain, affecting behavioral, motor, or other functions*? In line with previous results reported in the literature [194], we verified the role of the parietal lobe in the VR experience, as an indicator of the sense of presence. Our questionnaire confirmed the above. Both VR groups reported a high level of presence, although the one with the realistic appearance claimed higher levels. Moreover, through the questionnaire, participants reported low levels of spatial presence, especially in the VRI group and Spearman's correlation provided us with a positive correlation for this feature and the previous VR experience. We concluded that the more experienced users are in VR, the higher level of presence they can meet. Although participants didn't have a direct interaction with the VE, we noticed the sense of presence which means that interaction doesn't always play a significant role in the presence and immersion. It was not in our intention to urge participants to get involved physically with the VR task. Contrariwise, we wanted to determine possible differences in brain activity in several fixed environments and to examine if the design of a VE can also make a difference in this activity. However, we are aware that the possibility to move in the physical space could have given higher levels of presence and immersion, and maybe better results, but we tried to eliminate the noise in the recording of the EEG.

Of great interest is the presence of the theta state in the frontal region in both VR groups, indicating the cognitive effort and representing the attentional processing. To the best of our knowledge, such a result hasn't been reported in the literature before. However, the educational nature of the specific task possibly contributed to this activity, as the cognitive effort of an educational process might be bigger. Moreover, interesting is the dominance of the alpha band in the occipital lobe during VRI, indicating the difference between the two VEs in terms of visual attention mechanisms. It is also worth mentioning that through the questionnaire, we noticed that participants who experienced the normal VE (VR group) were more conscious of the real external stimuli and overall environment.

Having presented all our results, we need to mention that our study didn't involve any kind of physical interaction with the environment as our aim wasn't to urge participants to get actively involved in the VR task but to determine possible effects that several fixed environments can have in brain activity. The lack of activity was also present in the results of our questionnaire, as participants declared that they didn't feel

very active during both VR experiences. This, however, enhance our outcome as we know that it was the VR by itself that had all the influence.

Most of our participants, among the three groups, claimed that they would prefer to be trained via a VE rather than a PE for such an educational context. Given that we determined the impact of VR on the brain activity, and given that this impact is in line with the results of our questionnaire, there is strong evidence that VR has, for the time being, and under the existing technology, the potential to become a useful, efficient, reliable tool.

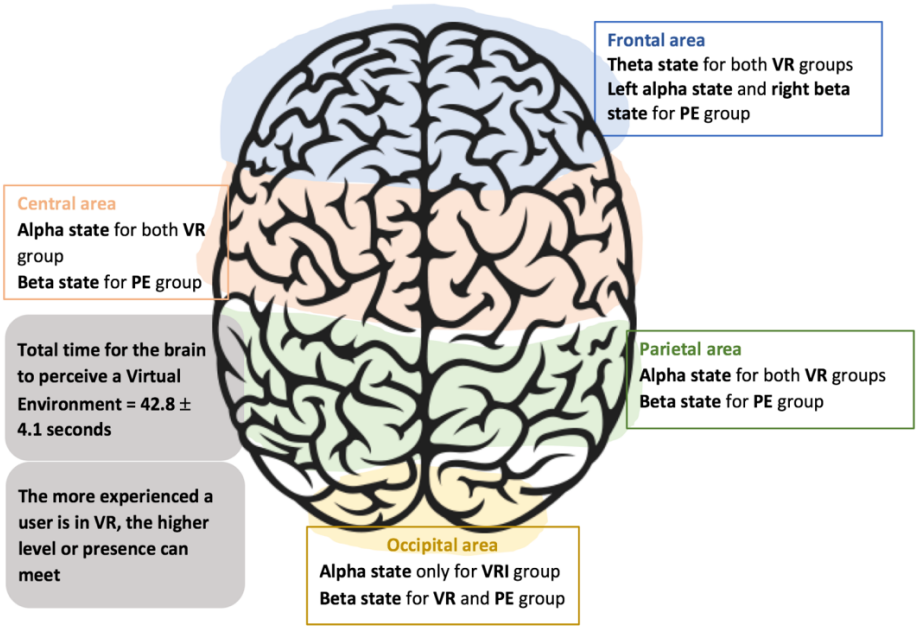


Figure 3.6 Summary of the results extracted from the EEG data and the questionnaire.

Figure 3.6 summarizes our results[123], combining EEG and the questionnaire. These results enhanced the SoA by :

- revealing some new brain states involved in VEs, like the Frontal theta state
- confirming existing ones, like the alpha waves in the parietal area and
- creating some links between the experience of the user and the sense of presence.

The outcome motivated us to proceed with our second study.

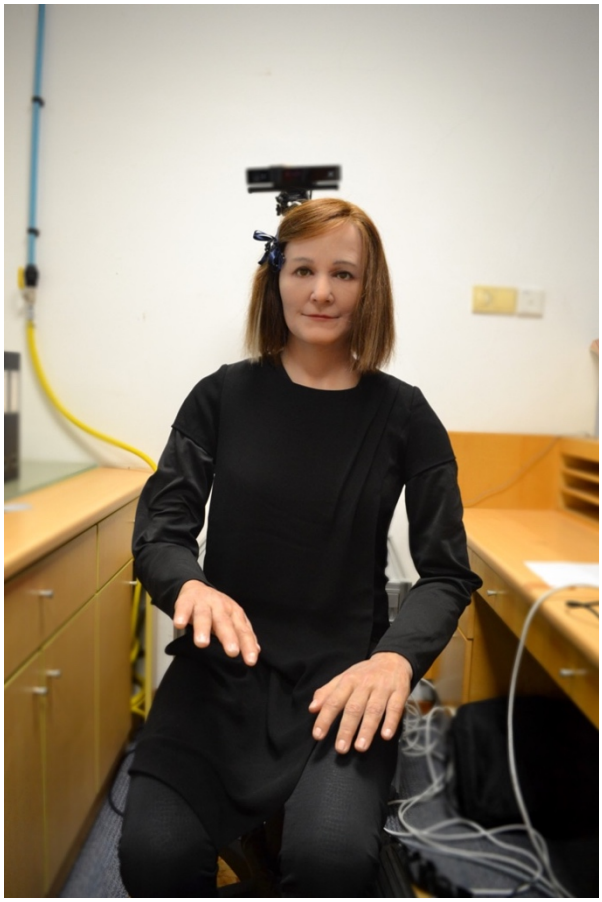
[1] Stavroulia, K. E., Christofi, M., Baka, E., Michael-Grigoriou, D., Magnenat-Thalmann, N., & Lanitis, A. (2019). Assessing the emotional impact of virtual reality-based teacher training. *The International Journal of Information and Learning Technology*. https:// doi.org/10.1108/IJILT-11-2018-0127

[2] Stavroulia KE, Baka E., Lanitis A., Magnenat – Thalmann N. (2018), Designing a virtual environment for teacher training: Enhancing presence and empathy. In proceedings of Computer Graphics International 2018 (CGI 2018), ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3208159.3208177

[3] Baka E., Stavroulia KE., Magnenat-Thalmann N., Lanitis A (2018), An EEG-based evaluation for Comparing the sense of presence between Virtual and Physical Environments. In proceedings of Computer Graphics International 2018 (CGI 2018), ACM, New York, USA, 10 pages. https://doi.org/10.1145/3208159.3208179

CHAPTER 4

HUMAN – ROBOT INTERACTION



"We are fascinated with robots because they are reflections of ourselves"

-Ken Goldberg, American Professor, artist, researcher

Human – Robot Interaction

4.1 Introduction

The outcome of our first research motivated us to explore more technological environments. We started wondering how the human brain would be affected by an environment with robots. Thus, we moved on to our second study, which was exclusively on HRI, and we tried to answer some first questions regarding the human's perception when interacting with a robot.

To develop empathic social robots that can persuade for their emotional awareness, it is important to consider human reactions derived from the interaction with them. The up-to-date research aims to develop robots or avatars that will be able to comprehend people. Thus, an interesting question is whether interacting with them could trigger human-like responses. To that end, enriching robots with a degree of emotional intelligence could lead to more efficient, meaningful, and natural human-robot interactions. Our purpose lies in examining the effects of human-humanoid interaction in humans' cognitive states and emotions, including the way the brain responds to such an interaction. The latter can give us an insight into the degree to which humans can perceive the difference of interacting with robots instead of other human beings. Most of the studies have tried to examine interactive tasks through observation, which means using video clips, or images [198, 201], highlighting the limitation of the physical interaction and the loss of the sense of embodiment.

To address the limitations of previous studies, we conducted an experiment, with three different types of interaction. Our main consideration is to examine if the brain can perceive the difference between a human and a robot that looks exactly like the human. In this chapter, we describe the experimental design of such an experiment, the methodology, and the results.

4.2 Experimental Design

During the experiment, volunteers were exposed to three different types of interaction under the same scenario. The first case (A) constitutes the control case and participants interacted with a neutral person. We chose to put the control case first to avoid any discomfort that might be caused by the interaction with the robot and to enhance the sense of familiarity during the whole process. The second case (B) concerns the human-robot interaction and participants had the opportunity to communicate with Nadine. Nadine is modeled on Prof. Nadia Thalmann, she has very natural-looking skin and hair and realistic hands, providing

a strong human-likeness. We chose to use the interaction with the real identical person as the third and last case (C) to examine if the brain can directly perceive the differences between the two last conditions. Figure 4.1 presents an example of a participant in the three cases. The whole experimental process was video and audio recorded with the consent of the participants.

The thematic areas of the discussion were pre-defined and guided by the people or the robot involved in the process, but the time of the interaction was up to the participants.

Figure 4.1 Participants during the three types of interaction. *LEFT*: Case A - Participant with a natural person, *MIDDLE*: Case B - Participants interacting with Nadine, *RIGHT*: Case C - Participant discussing with Prof. Nadia Magnenat Thalmann

Table 4.1 presents the chosen topics of discussion and the expected emotions for each condition. The discrete emotions are chosen based on the questionnaire we used.

Table 4.1 Thematic areas used for the facilitation of the discussion during the three interaction

Question Topic	Example	Probable emotions triggered
Introduction	"Hi, my name is Nadine. What is your name?"	Joy, Fear, Interest, Nervous
Profession	"What do you do for a living?"	Joy, Calm, Inspired
Family and relationships	"Are you married, or do you have a partner?"	Confident, Ashamed, Nervous
Hobbies and recreation	"What are your hobbies?"	Calm, Joy, Inspired
Religious Views	"Do you believe in the existence of God?"	Inspired, Fear, Ashamed, Calm
Miscellaneous	"Do you like this experiment?"	Interested, Nervous

For the validation and the accuracy of our experiment, that is to properly examine the effect of such interactions, we used EEG recording to capture the brain activity, an audio recorder to capture voice features, and psychometric measures through a questionnaire that subjectively examined the emotional states of the participants.

4.2.1 Participants

This study acted as a preliminary study and thus, the number of our participants was limited; though enough to have a decent statistical validity. So, twelve healthy adults, aged from 20 to 35, participated voluntarily in this study. The study took place in the Institute for Media and Innovation (IMI) at the Nanyang Technological University (NTU) as the humanoid robot Nadine was currently situated there. We tried to ensure no previous experience of the participants with robots to avoid any bias in our results. A form of consent, based on the NTU requirements, was signed by all the subjects before the onset of the experiment. None of them mentioned any sign of discomfort.

4.3 Data Collection and Analysis

4.3.1 EEG recordings and analysis

EEG signals were recorded and amplified using a NuAmps amplifier (https://compumedicsneuroscan.com/applications/eeg/). 34 electrodes were attached on a Quick-Cap according to 10-20 system at the locations Fp, F, FT, FC, C, T, TP, CP, P, PO, O. Curry 8 X was used for the data acquisition and the online processing with a sample rate 1kHz per channel.

The analysis of the EEG data and the processing of the signal were carried out in MATLAB. All data were carefully checked for artifacts, like eye blinks or head/body movements. Fast Fourier Transform was applied to the signal to transport it to the frequency domain and then the power spectra was calculated. The analysis was conducted in two setups. The first one consists of 5 Regions of Interest (ROI) examining five brain areas: Prefrontal (Fp), Frontal (F), Parietal (P), Temporal (T), and Occipital (O) whereas the second one analyses the same areas but for each hemisphere (10 ROIs). Figure 4.2 illustrates the examined ROIs. Brain rhythms that we are mainly interested in are theta (3-7 Hz), alpha (8-12), beta (13-30 Hz), and low gamma (30-42 Hz).

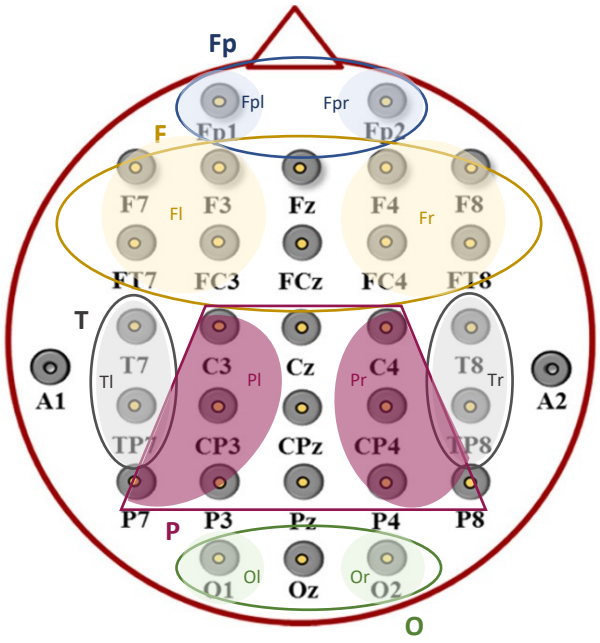


Figure 4.2 Regions of Interest (ROIs) used for this study. 32 EEG electrodes were used, with the help of a QuickCap, attached according to the 10-20 system. We examined two groups of ROIs. The first corresponds to the five brain areas (Prefrontal – Fp, Frontal – F, Temporal – T, Parietal – P, Occipital – O) and the second to the five brain areas for each hemisphere (Fpleft – Fpl and Fpright – Fpr, Fl and Fr, Pl and Pr, Tl and Tr, Ol and Or)

Based on previous studies, we will focus our research in the frontal and parietal areas [193]and we will use the occipital region to investigate the recruitment of visual attention mechanisms, as before. We will also examine the possible existence of frontal theta oscillations that some studies have already noticed during HRI [201].

The experiment was conducted in a sitting position so that we could minimize as much as possible the noise and probable distraction from the cables or other features.

4.3.2 Dialog and audio analysis

The language was understood by Nadine using Google Cloud Speech-to-Text transcription and the whole conversation was held in English. The audio signal was processed using Praat software, extracting the features of pitch and intensity. The duration of each interaction was also assessed. The human participants were from different cultural backgrounds with unique styles of conversation. The thematic areas though were specific, no matter the choice of each interviewer. On the other hand, Nadine was operated in two modes: *control* and *free* mode.

The social robot Nadine's architecture

Our humanoid in the *control* mode used the Wizard-of-Oz technique and the questions in the interviews were asked in an orderly fashion. The participants had to respond to each of those questions, followed by the free *mode*, where the participant was asked to ask for anything. In the *free* mode, the answers are based on a chatbot with the architecture in Figure 4.3. In our experiment, the episodic memory portion is ignored, since the participant was unfamiliar to Nadine.

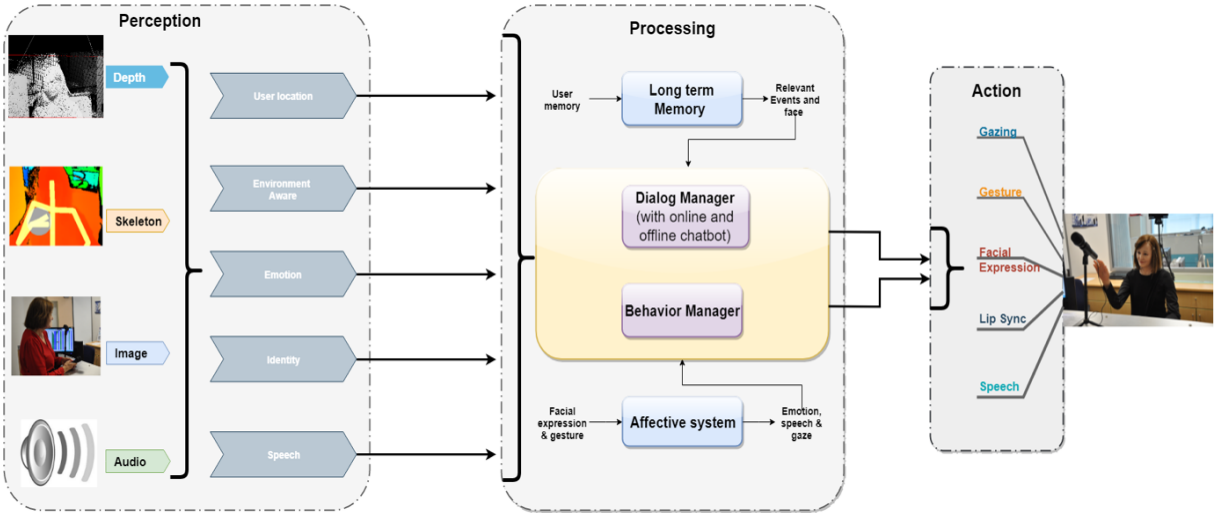


Figure 4.3 Nadine's architecture [116]. A social robot (Nadine) must mimic humans in all possible scenarios and behaviors. Like any human, she should be able to process all information about the environment, user, and context to come up with appropriate correct responses and reactions. Nadine's perception layer helps her to collect information about the environment such as where is the human located, capturing human's face images, capturing images of the environment, and getting speech of the user. The processing layer is the core module of Nadine that receives all results from the perception layer about the environment and user to act upon them. Each perception layer output is processed in this layer taking into account the customization done to come up with appropriate verbal or nonverbal responses. Verbal responses are spoken by Nadine and nonverbal responses are shown by Nadine in the form of gesture, facial expression, and lip-sync with a verbal response.

In our experiment, the episodic memory portion is ignored, since the participant was unfamiliar to Nadine. Therefore, once the speech of the participant is converted to text, it is sent to the chatbot. If the chatbot does not have an appropriate response, it is looked up online. If the online results are not available, a generic default response is given.

4.3.3 Psychometric data

To provide our results with a higher validity, a reliable and validated questionnaire, including closed-ended Likert-scale questions, was used. The questionnaire consisted of questions regarding participants'

demographic data and mood states for each condition. The emotions scale was based on the Positive and Negative Affect Schedule, which comprises two scales: one measuring positive affect and the other measuring negative.

4.3.4 Statistical analysis

Statistical analysis was carried out for the variables of the EEG and the questionnaire, through SPSS. We conducted Repeated Measures ANOVAs and followed up statistically significant results with the Bonferonni post hoc tests. When the data did not meet the sphericity requirement, the corresponding non-parametric Friedman test was used and the corresponding post hoc tests (Connover's Post Hoc Comparisons - Conditions test) validated the statistical significance.

4.4 Human Perception during Human-Humanoid Interaction and its Effects in Human Cognitive and Emotional States

Having verified that VR can indeed affect several functions, we were motivated to go a step further and examine human perception towards another framework; the one of human-robot interaction. We were interested to see if the interaction with a robot can trigger human-like responses and how such an outcome could lead to a more efficient and natural HRI. Hence, our purpose lies in examining the effects of human-humanoid interaction in humans' cognitive and emotional states and the degree to which a human can perceive the difference between a robot and another human being.

4.4.1 Brain activity during human-humanoid interaction

As we have mentioned before, we were focused on five brain areas. Starting from the prefrontal cortex, we noticed a general alpha rhythm, the same for all three conditions with no significance difference ($F(2, 22) = .43$, $p=0.654$). Prefrontal cortex is fully associated with the personality, planning of complex and social behaviors, decision making and in general with the orientation of our behavior in line with our goals and values [229].

For the frontal cortex, we observed a significant difference between the two human cases and the one of the robot. In both H-H interaction (A and C) we noticed a high alpha state, around 12 to 12.2 Hz whereas, during the interaction with Nadine, theta oscillations were observed (7.8 ± 0.6 Hz) ($p=0.002$). In this case, sphericity requirements were not met, and non-parametric tests were used (Friedman Test and Connover's Post Hoc tests). As we mentioned before, frontal theta oscillations are associated with focused attentional

processing and they increase as a task becomes more demanding [230]. Thus, the outcome can be attributed to the participants' bigger cognitive effort to get focused while interacting with Nadine. Moreover, this result is in line with other studies that have noted the existence of theta oscillations when a human interacts with a humanoid robot [201]. The case however is not the same when interacting with a mechanical one.

Our results in the parietal cortex complement the above. We noticed beta oscillations in both H-H interactions (13.4 ± 3.8 Hz for the A case and 14.7 ± 5.5 Hz for the C case) whereas in B case we found dominance of the alpha state (10.8 ± 2.6 Hz).

In the temporal cortex, we noticed the same pattern as in the parietal area. Both H-H interactions were characterized by beta oscillations (A: 14.2 ± 4.8 Hz, C: 21.1 ± 5.3 Hz) whereas the HRI by alpha (B: 12.2 ± 2.8 Hz) with $p < 0.001$, indicating a high statistically significance. However, the post hoc analyses exhibited that case C differed from case A and B ($p = .007$) and $p<.001$ respectively). The temporal lobe, in general, is associated with processing of auditory information [230]. The presence of the alpha band during HRI may indicate that participants put a higher effort to decipher Nadine's speech compared to the human's way of talking they are used to.

Regarding the occipital region, the results were as expected. The analysis exhibited statistically significant differences for the three cases($F(2, 22) = 8.06$, $p=.002$). The post hoc analysis exhibited that Nadine's case (B) presented an alpha rhythm (9.7 ± 2.3 Hz) that differed from both A and C cases (A: 14.5 ± 5.7 Hz, C: 19.4 ± 7.6 Hz), ($p=.004$ and $p = .090$, respectively). This result proves that the brain can completely visually understand the difference between a human and a robot, whatever appearance the latter may have. However, we can also notice the higher value of the C case, which reveals a bigger familiarity compared to the A case. That may be explained by the fact that participants had first seen Nadine and then Nadia, so they already had created a memory image of this appearance.

Figure 4.4 depicts the results described above for the five brain areas.

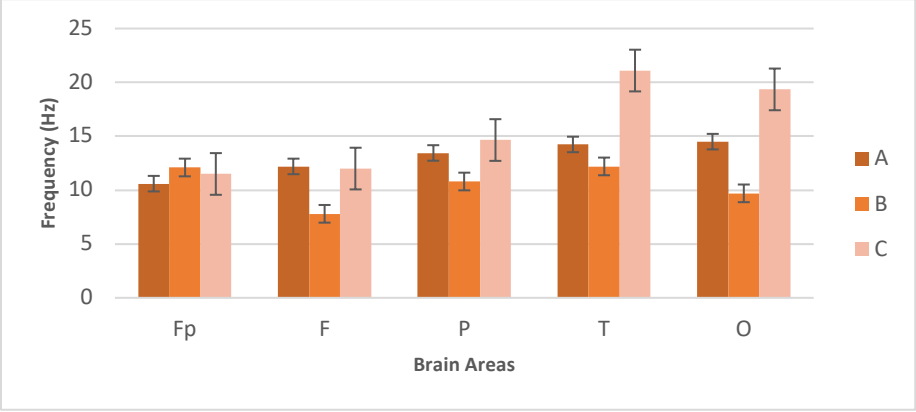


Figure 4.4 Power spectra observed in each of the 5 ROIs, in response to each case. The 5 ROIs have been depicted in Fig. 4.2. Power spectra was calculated as $PS = abs(filteredsignal)^{\wedge}2$.

For all these areas, we also examined possible differences between the two hemispheres. Figure 4.5 presents the differences among the three conditions for each hemisphere. No significant differences were noticed for the prefrontal cortex in the three conditions. For the frontal cortex though, we saw that the left hemisphere had a bigger activity during the C case (12.3 Hz ± 3.9 Hz) and only the latter had a significant difference with HRI (case B). The right hemisphere though presented higher activity during the case A but both H-H interactions were significant different from case B. For the parietal cortex, the left part of the brain didn't present any difference whereas for the right part, only the A case shows a significant difference with case B. For the temporal cortex, both in right and left hemispheres the differences are statistically significant. Lastly, for the occipital region, we have the same result in both hemispheres, meaning that cases A and B present a significant difference compared to case C.

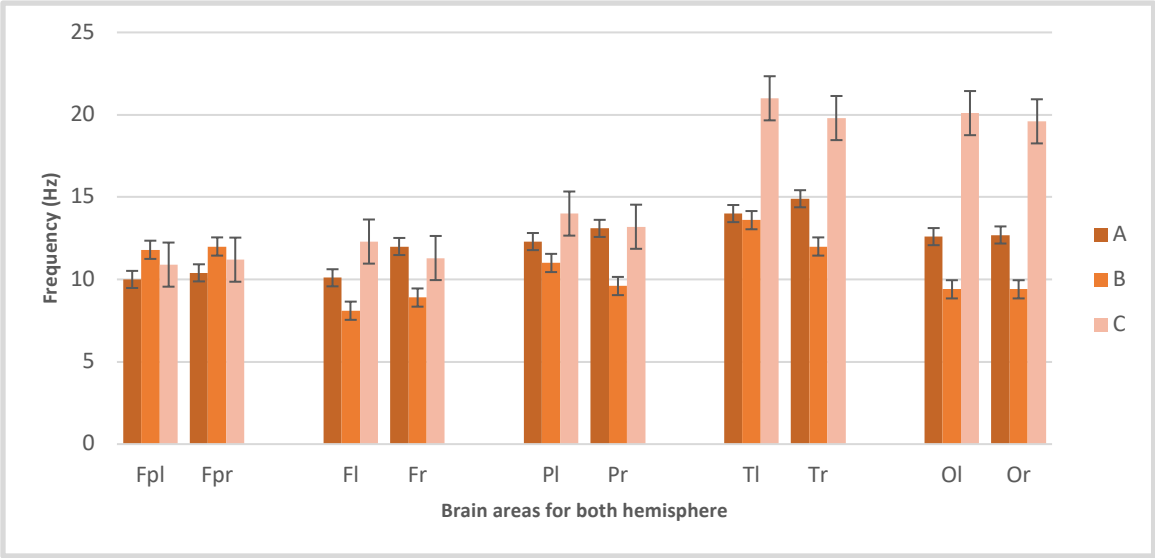


Figure 4.5 Mean of frequencies for each brain area for both hemispheres in three conditions

The second, and maybe a bit more interesting question, is if each case has a different effect in the two hemispheres. Thus, we run statistical tests for the three cases separately and we concluded to the results depicted in Figure 4.6. For the case A, we noticed a slightly bigger activity for the right hemisphere, but ANOVA tests revealed no significant differences between the two hemispheres. For the case B, we noticed exactly the opposite, with the left hemisphere presenting higher activity. Parametric post hoc tests showed significant validity only for the frontal and parietal areas. For case C, we noticed again higher values of frequencies in the left hemisphere, but ANOVA tests revealed no statistical significance.

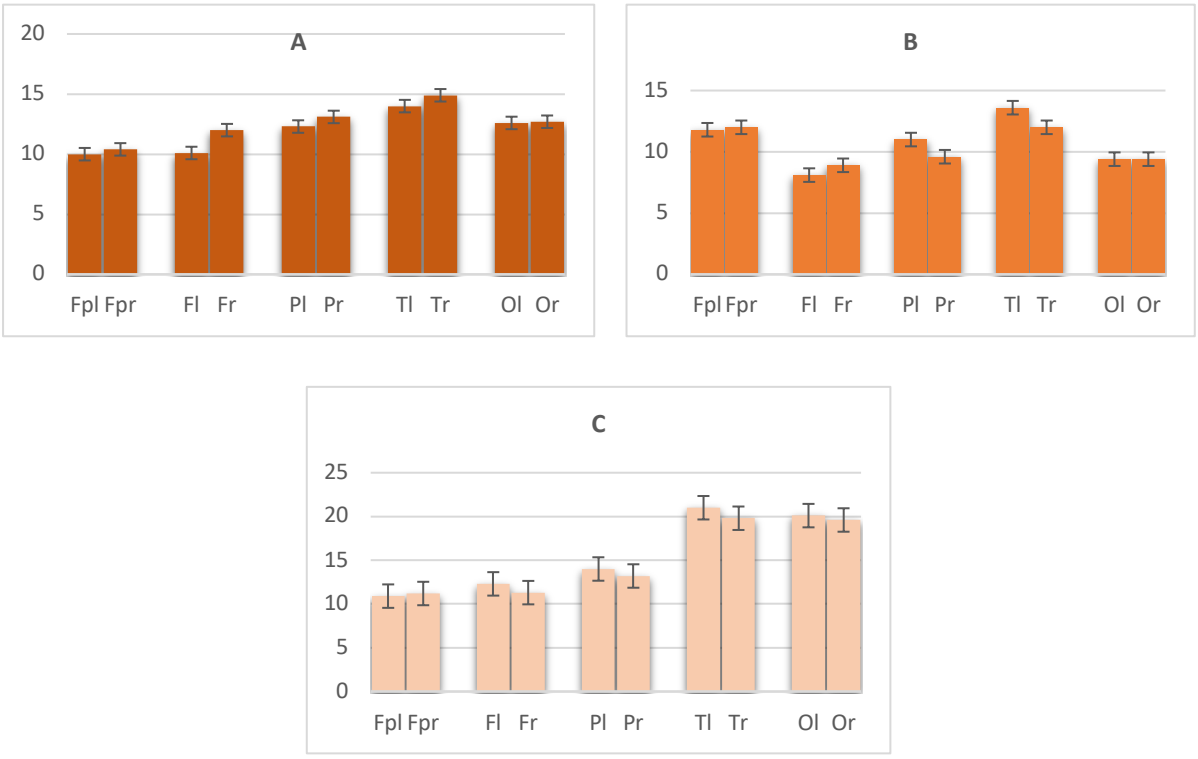


Figure 4.6 Differences in frequencies of each case for the two hemispheres

4.4.2 Audio data and human perception between HH and HR Interactions

From what EEG data revealed, the main difference in human perception when interacting with a human and an identical human-like robot is in the visual perception. However, this is normal and expected as people are not yet familiar with such kind of technology and interactions, not with physically present robots. The analysis of the audio data complemented the above.

We observed the higher pitch of the voice (= 133 Hz) during the humanoid conversation. Although it may be inconclusive due to our small sample size, it seems promising and motivating to explore the hypothesis that people speak in a higher pitch when interacting with humanoids. This may reveal a lower level of comfort and a higher level of nervousness during such interaction. It can also be combined with the outcome of the EEG analysis in the temporal cortex, where the presence of alpha band was noticed only for the case of Nadine, indicating the higher cognitive effort of the participants to understand Nadine's voice.

Regarding the intensity, we found that participants had a louder voice when interacting with humans, with a significant difference from the HRI ($p = .011$ for case A and $p = .003$ for case C). We also examined each thematic area of the discussion separately and we found a peak in the intensity during the question regarding

the belief in the existence of God. This can be due to the emotional nature of the question or possible discomfort.

There was also a noticeable delay in the speech during the HRI. However, the duration of the conversation was longer during the case C. In figure 4.7 we can find the comparison between the three audio features described above.

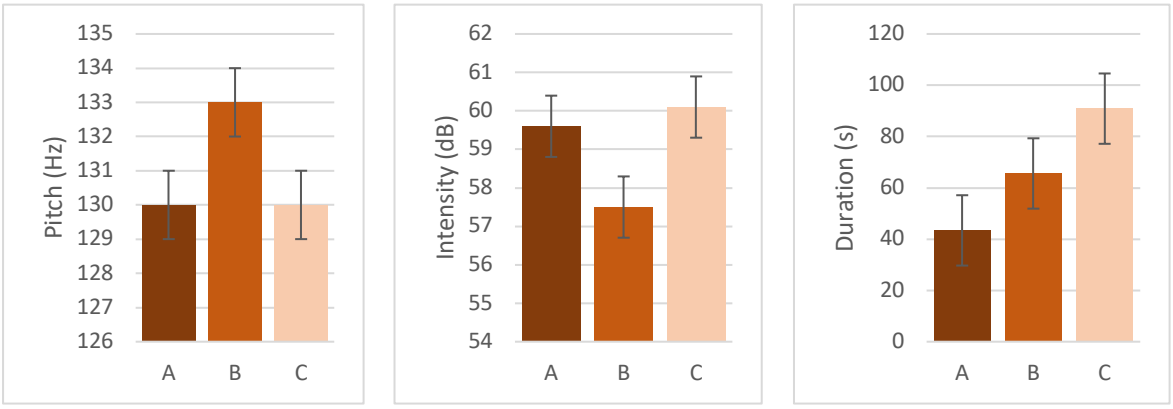


Figure 4.7 Mean values for Pitch, Intensity, and duration of the interaction for each condition, derived from the audio analysis

4.4.3 Differences in emotions and motivation when interacting with a human and an identical robot

Based on the EEG data, the existence of the alpha state in the parietal cortex during the HRI led us to the conclusion that during such an interaction, humans are unintentionally more concentrated on their tasks. We can attribute that to the existence of new, non-familiar elements people are forced to face. It is the same result we retrieved from the exposure of people to VR environments, as we presented in our previous study. Thus, we can claim that people unintentionally tend to be more concentrated when they are exposed to environments that they are not familiar with.

The longer duration of the conversation in case C comes to verify the results we acquired from our questionnaire where participants showed a higher amount of inspiration and concentration during that case. We can also see that the duration with the neutral human was lower than the humanoid which is also verified by the questionnaire results. Moreover, we can note that the conversation with the humanoid had no interruptions by the participant when compared to humans. This highlights the limitation of the humanoids of the present generation which lack a fully natural conversational ability.

Based on our questionnaire we noticed no significant negative emotions during the whole procedure. The most dominant emotions were Interest, Inspiration and Confidence which reveal also a sense of motivation.

It is also worth mentioning that the state of inspiration appears from case B and on, with an ascending value, meaning that Nadine triggered this reaction to the participants who kept being inspired for the rest of the procedure. We may also justify this by the fact that they found the similarity in the appearance interesting and motivating. At this point, we could have expected to find also a negative emotion, like scared or anxious, due to the high degree of human-likeness and the UVH. The fact that no negative emotions were noticed reveals that the nature of the interaction (the scenario or even the purpose) can affect the UVH and the human perception.

The state of the interest is higher in case B which is normal if we assume that people are not yet so used to the existence of robots and they don't often have the chance to interact with them. The state of concentration was increasing along the process. Lastly, the participants felt significantly active only in cases B and C. Figure 4.8 presents the results of the questionnaire for the three cases.

However, in the question of who was the most comfortable to discuss with, participants preferred the human existence, voting equally the neutral person and Professor Nadia Thalmann.

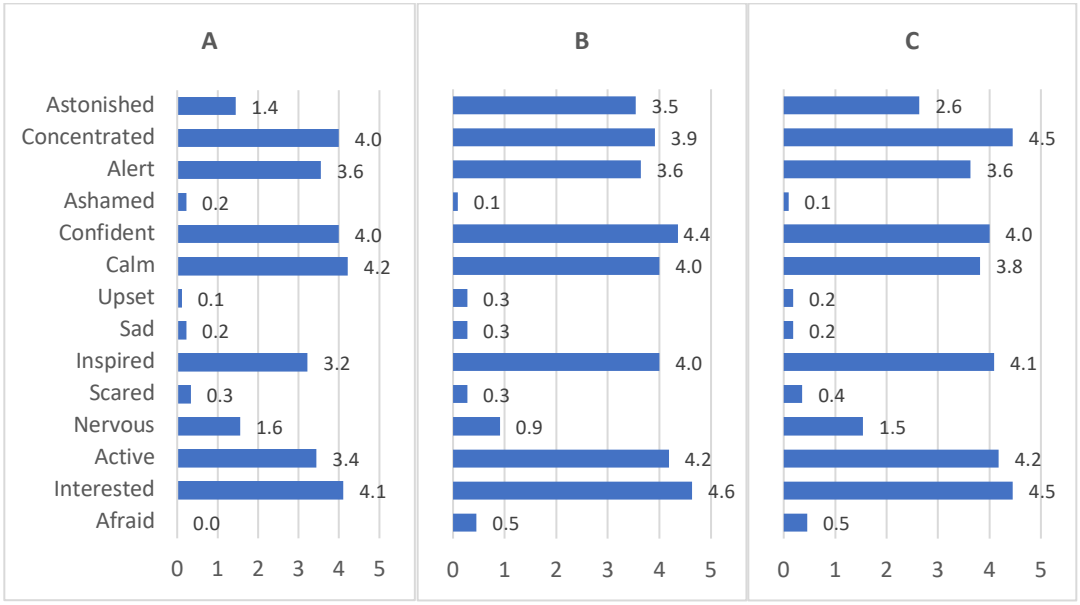


Figure 4.8 Participants' emotional states for each condition. The questionnaire was distributed at the end of the whole procedure.

Summing up, we noticed that there is a motivation enhancement throughout our process, with no observation of negative feelings. Participants, in the beginning, were presented as calm, confident, and concentrated and while interacting with Nadine new states appeared like inspiration, activation, and interest. The values of all the emotional states were ascending, revealing the high level of motivation. The increasing duration of speech in the three scenarios verifies the same.

4.5 Summary and Discussion

In this preliminary study, we investigated the human perception during human-humanoid interaction and its effects on cognitive and emotional states. We have used three cases where people face three different types of interaction. To support our research, we used EEG and audio recordings as well as a questionnaire to complement the psychometric data.

With this study, we answer our second research question: *Is there a difference in the perception of a human and an identical human-like robot?* Our results revealed a difference in the perception of a human and an identical human-like robot mainly in visual perception. We consider that people are not yet familiar with physically present robots, so this is normal. This difference was uncovered by the existence of the alpha state in the occipital cortex, which proves the activation of visual attention mechanisms compared to the human-human interaction where we noticed the existence of beta states and consequently, the sense of familiarity. The latter was more enhanced in the third case, where the human interacted with professor Nadia Magnenat-Thalmann and thus, we are not sure if this familiarity is a result of previous interaction with the humanoid.

Regarding brain activity, we found some very interesting results. In the prefrontal cortex, we found no difference between the three cases, all of them synchronized in alpha brain state. However, we noticed frontal theta oscillation in case B, Human-Nadine interaction, with a clear difference from the other two conditions where we saw the dominance of the alpha state. This comes in line with previous studies [201] that have noticed the existence of theta oscillations when a human interacts with a humanoid but not with a mechanical robot.

The same result was observed in the parietal cortex where, only in the human-robot interaction, we noticed the existence of the alpha state, which is associated with the perception process. In general, we conclude that during such an interaction, humans are unintentionally more concentrated on their tasks and we can attribute that to the existence of new, non-familiar elements people are forced to face. This has also been supported by the results of the questionnaire used, where emotions related to motivational states are developed during the human-robot interaction. This is in line with our previous studies regarding the comparison between VR and physical environments where it was found that people tend to unintentionally be more concentrated on an environment they are not familiar with [231].

Of great interest is the result of the temporal area, which is linked with the processing of auditory information, and is verified by the outcome of the audio analysis. We noticed the presence of the alpha power only in case B which can be attributed to a higher cognitive effort participant made to understand Nadine′s voice. This can also be correlated with the higher pitch of the voice participants presented during

their interaction with Nadine. It could also be a result of nervousness, but the results of the questionnaire didn't reveal any negative emotions. Figure 4.9 summarizes the results of the brain activity and possible connections with the audio signal.

Completing the answer to the second research question, we need to reply to the subsection of it: *Is there a difference in emotions and motivation when simply interacting with a human and an identical robot?* Thus, we concluded that there is a motivation enhancement throughout our process, with no observation of negative feelings. In the beginning, participants were presented as calm, confident, and concentrated and while interacting with Nadine new states appeared like inspired, active, and interested. The values of all the emotional states were ascending, revealing the high level of motivation. The increasing duration of speech in the three scenarios verifies the same.

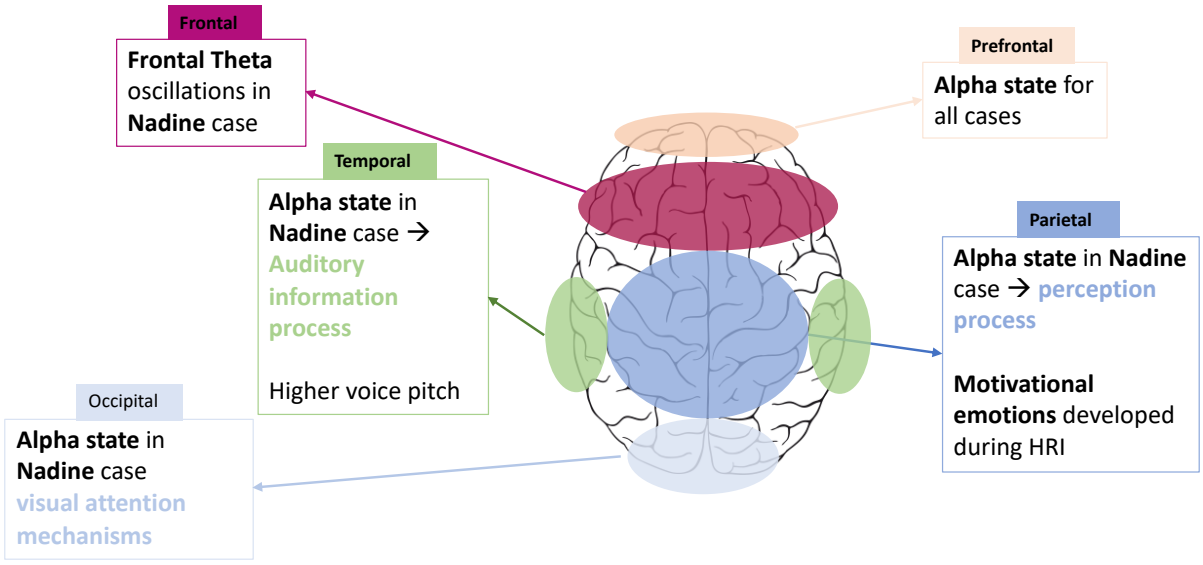


Figure 4.9 Summary of the results

To sum up, we remind that the purpose of this study was to provide a first glance at this innovative approach of human-robot interaction, examining human cognitive states when interacting with a robot and a human that look alike. Robots and virtual characters are increasingly becoming ubiquitous in our daily lives. Therefore, it is paramount to study our behaviour and emotions to these technological transitions to aid in their human development and to enhance their applications in several domains like education, rehabilitation, or even entertainment.

Our work[1] contributed to the SoA as it was the first study to compare a humanoid with an identical human and assess the human perception, concluding that the human brain does understand visually and auditorily the difference but the level of concentration remains higher during HRI.

[1] Baka E., Vishwanath A., Mishra N., Vleioras G, Magnenat Thalmann N.(2019), "Am I talking to a Human or a Robot?": A preliminary study of Human's perception during Human – Humanoid interaction and its effects in cognitive and emotional states. M. Gavrilova et al. (Eds.): CGI 2019, LNCS 11542, pp. 240–252, 2019. https://doi.org/10.1007/978-3-030-22514-8_20

CHAPTER 5

HUMAN – NONHUMAN INTERACTION

"I do not fear computers. I fear the lack of them."

- Isaac Asimov, American writer and professor

Human – Nonhuman Interaction

5.1 Introduction

Robotics and virtual agents can improve the accessibility of various content. As we have mentioned earlier, what we need is to find the key point where humans socially approve and accept social agents and social agents have as a principle the human needs. With the current COVID-19 pandemic making our social life difficult and increasing the level of stress and vulnerability among populations [232], the development of more efficient technology-assisted interventions is crucial. Additionally, the existence of nonhumans agents that can make humans feel comfortable and more motivated can make a difference. Although different kinds of agents have been used to contribute to several domains, like education, health, entertainment, both in virtual and physical environments, the digital human or robot that will make a human feel as comfortable as interacting with another human has not been reported yet.

Inspired by the results of the previously described experiment, we decided to go a step further, delving deeper into the features and the nature of H-H and H-NH interaction. Thus, we designed an experiment where our main purpose is to use human complete behavior to reveal the humans' needs towards technology and to possibly find a way to suggest an improvement. We performed in-depth documentation, analysis, and comparison between the natural human-human and the human-nonhuman interaction, to find the gaps between them revealing humans' needs and what can affect the efficiency, fluidity, and naturalness of an interaction. We also examined possible correlations among the extracted features to examine how and if our reactions are related. Moreover, we tested HRI under several roles to examine possible changes in humans' perception and acceptance towards robots. To the best of our knowledge, such multidisciplinary and in-depth documentation, analysis, and comparison between H-H and H-NH interactions has not been described in the current literature yet.

Thus, this chapter is divided into two parts. Subchapters 5.2 and 5.3 describe the first and the second experiment respectively, providing detailed information on the experimental design, the methodology, the results, and the conclusions. Finally, 5.4 discusses the overall outcome of this work, pointing out the contribution to the existing SoA.

5.2 Human Behaviors and Reactions during H-NH Interactions

5.2.1 Experimental design

To serve our purpose, our experiment consists of two parts. The *first part* includes three different types of interaction under the same scenario.

- Interaction of a human with Nadine humanoid social robot (N)
- Interaction of a human with a virtual human (VH)
- Interaction of a human with another human (H)

The scenario simulates the first phase of a job interview where participants had to answer several predefined questions and present themselves. All participants took part in all three different types of interactions. Figure 5.1 shows an example of a participant interacting with the virtual human Nicole and presents the two nonhuman agents used. Given that participants had to face all three interactions, the sequence was different for each one of them to exclude any possible bias during our feature extraction (like preparation of the answers or familiarity with the nonhuman agents). The whole procedure took place in an isolated room under identical circumstances and with no external noises. The experiment was in collaboration with the Institute for Media and Innovation (IMI) at the Nanyang Technological University (NTU) and it took place in NTU where we had full access to Nadine social robot and the virtual human Nicole. Our first plan, based on our results from our first experiment regarding the physical and virtual environments, was to use Nicole in a VR setting. However, we decided to keep the experiment as simple as possible to simulate conditions that are more accessible to everybody.

Figure 5.1 LEFT: case VH where the participant interacted with Nicole, the virtual human. RIGHT: the nonhuman agents used in our experiment: Left: Nicole the virtual human, Right: Nadine the social robot

To validate our study, we followed a multimodal approach, and we used the following human modalities:

- EEG to capture brain activity
- An audio recorder to capture voice and speech features
- Kinect to record and examine body skeleton movements
- EMG to capture the activity of specifically selected muscles
- A questionnaire (Panas X) to assess humans' emotions, mood states, and the overall experience of the interactions.
- A Negative Attitude Towards Robot scale (NARS) to assess users' overall attitude towards robots

Figure 5.2 presents the modalities and the extracted features of our work.

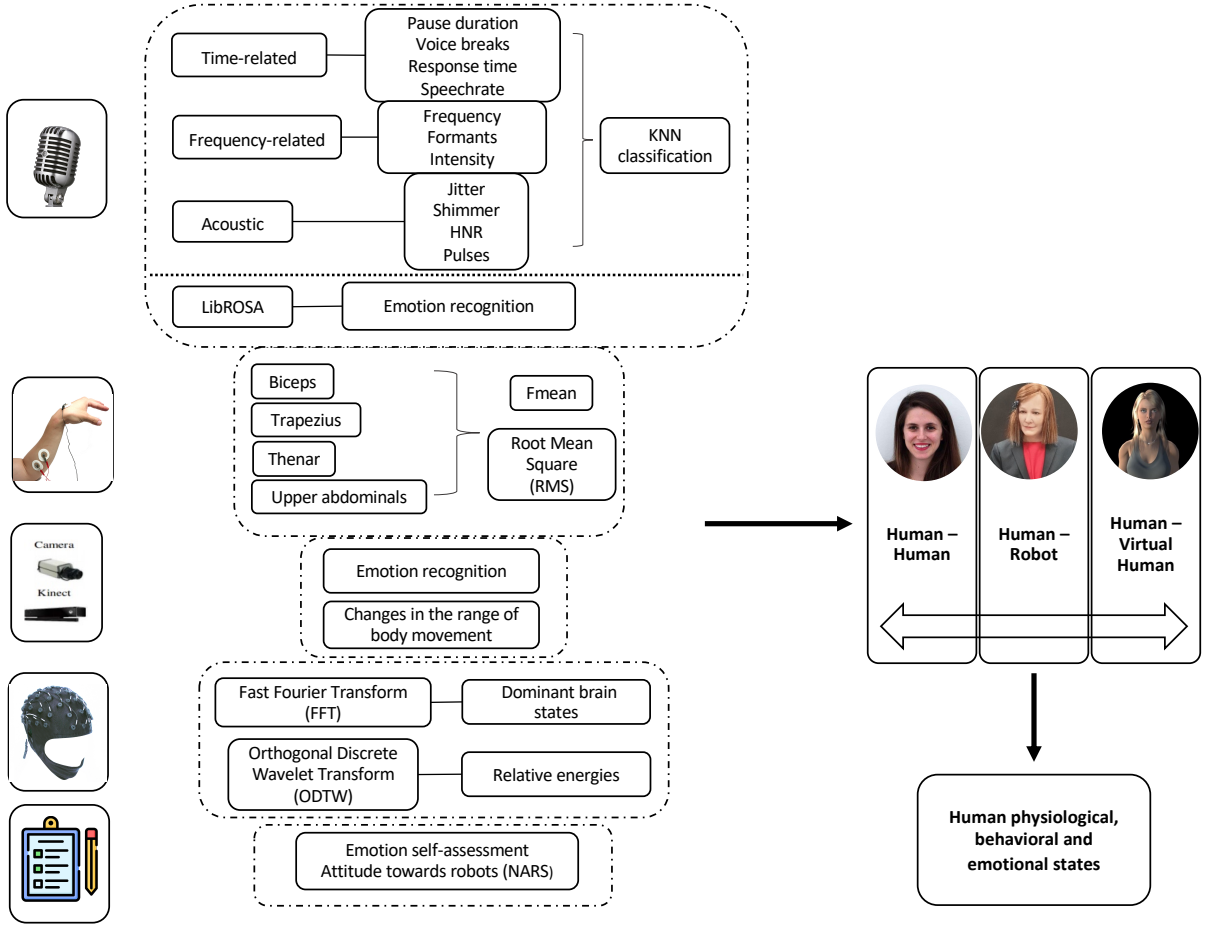


Figure 5.2 The setup of our work

All three interactions had the same job interview scenario and the questions asked were predefined. However, the content and the order of the questions were different for each case so that participants

wouldn't get directly familiar and used to them. Table 5.1 presents the thematic areas and the relevant questions used for the case of Nadine the robot.

All interactions were held in English. The overall duration of the experiment for each participant was around one hour, including the time needed for the setup of the devices, the explanation of the procedure, the completion of the questionnaire, and the consent form. The expected outcome was to obtain global information on how humans react during interactions with digital humans and robots and compare the latter with a natural H-H interaction under the same scenario.

Trying to avoid any latencies or lack of synchronization due to the simultaneous multimodal recordings, we performed a fast calibration for all the modalities. We used as a base the EEG signal and participants were asked to relax in their position with their eyes open and closed periodically. Then, they were asked to move their hands alternately. The procedure lasted for one minute. In this way, we ensured that signals were influenced by the human reactions and were synchronized.

Table 5.1 The thematic areas used for the job interview scenario and the questions for each one of them as posed for the case of Nadine robot.

Job Interview scenario		
Thematic Areas	**Questions**	**Keyword**
Introduction	"How could you describe yourself?"	Description
Hobbies	"What do you like to do outside of work?"	Hobbies
Personal info/ Emotion triggering	"Why do you think we should hire you?"	Suitability
	"What is your greater weakness?"	Weakness
	"What do you consider to be your strength?"	Strength
	"How do you handle stress and pressure?"	Stress
Previous work experience	"What is your greatest professional achievement?"	Prof_Ach
	"Tell me about a challenge or conflict you have faced at work and how you dealt with it?"	Challenge
Current work requirements	"What type of work environment do you prefer?"	Work_env
	"What are your salary requirements?"	Salary
Future expectations	"Where and how do you see yourself in five years from now?"	Yourself_5

Figure 5.3 summarizes the steps of our research protocol.

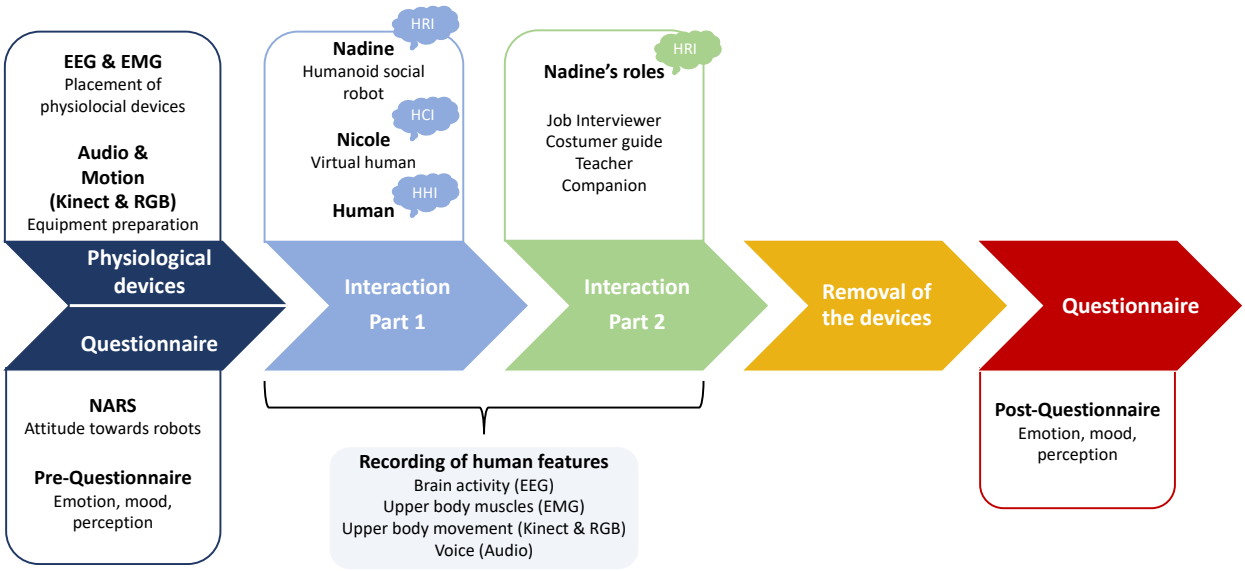


Figure 5.3 The flow of our research protocol

5.2.2 Participants

Forty individuals participated in our study. They represented a broad range of ages (20 to 65 years old) and two ethnicities (Asians and Europeans). Table 5.2 provides a detailed description of our sample's demographic data. Although gender and ethnicity were not balanced, we examined possible differences in the way of communication and interaction, acknowledging only high statistical differences. We assured that participants had no previous experience with robots, digital humans or any similar technology. Based on the requirements of the University's Institutional Review Board, a consent form followed by a detailed explanation of the experiment was signed by each participant before each experiment. Each participant received a small compensation for contributing and helping to our research.

Table 5.2 Demographic data of the study sample

Participants' characteristics		**Total Number (Percentage %)**
Role	Student	19 (48%)
	Employee	19 (48%)
	Retiring	2 (4%)
Gender	Female	13 (33%)
	Male	27 (67%)
Ethnicity	European	8 (20%)
	Asian	32 (80%)
Age	20 – 30	29 (72%)
	31 – 40	7 (18%)
	>40	4 (10%)

5.2.3 Data acquisition and analysis

EEG recordings and analysis

EEG signals were recorded and amplified using a NuAmps amplifier (https://compumedicsneuroscan.com/applications/eeg/). Given our research interest, we examined specific brain areas, so we used in total 23 channels. These channels were attached on a Quick-cap according to 10-20 system over the locations Fp, F, FC, T, CP, P, O. Specifically, electrodes were placed over the positions Fp1, Fp2, F7, F3, Fz, F4, F8, FC3, FCz, FC4, T7, T8, TP7, TP8, CP3, CPz, CP4, P3, Pz, P4, O1, Oz, O2, covering the Prefrontal (PF), Frontal (F), CentroParietal (CP), Parietal (P), Temporal (T) and Occipital (O) cortexes. For this part of the experiment, we used 6 ROIs. Two more reference electrodes were attached to the earlobes. A ground electrode was also placed on the Cz position. For the data acquisition, Curry 8 X was used, electrode impedances were kept lower than 2 kΩs and the sample rate was 1kHz.

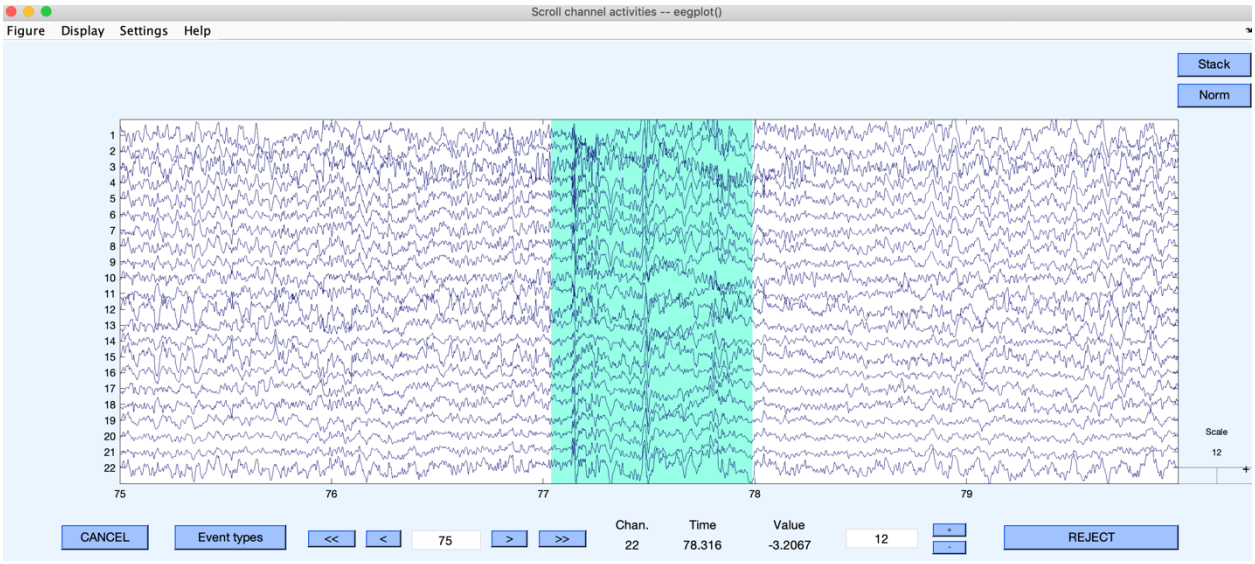


Figure 5.4 Example of the visual manual inspection conducted in EEGLab.

The pre-processing steps were performed in Matlab, partially with the help of the EEGLAB graphic user interface [233]. We applied a high-pass filter with a cut-off frequency at 1 Hz to remove low-frequency signals and a notch filter centered on 50 Hz to eliminate the industrial noise. Both filters were Butterworth digital filters of 3rd order. Then, an Independent Component Analysis (ICA) was performed to reveal and reject artifactual resources, like eye blinks, muscle artifacts, bad electrode placements, linear trends, and high frequency noise. A further visual manual inspection was conducted to reject short data segments that were contaminated with noise, as shown in Figure 5.4. The final filtered signal was selected as an input to the synchronization analysis algorithm. This final signal for each interaction of each participant had a

minimum duration of 20 seconds, as it has been proven that this time interval is sufficient for the extraction of the synchronization degree [234].

The synchronization analysis of the EEG data was performed on the entire EEG activity. It aims at the extraction of the activity for the five frequency bands (delta, theta, alpha, beta, gamma) for each electrode and its relative energy contribution. For that purpose, we used wavelets which are mathematic oscillatory tools fitting the time-frequency analysis of non-stationary data [235]. The first step was the selection of the appropriate mother wavelet, which specified the basic shape of the wavelet. Then, the entire wavelet family was subjected to scaling and translation to extract both frequency and time-dependent components respectively. As mother wavelet, the family of 5^{th} bi-orthogonal wavelets was selected due to its resemblance with the common EEG waveforms as well as its mathematical properties (symmetry, semi-orthogonaliry, maximum time-frequency resolution, and smoothness) [234, 235]. Therefore, phase distortion and discontinuity effects are avoided [235, 236]. The epochs of our EEG continuous data were divided into windows of 128 ms duration. The first 150 windows were further analyzed. The Orthogonal Discrete Wavelet Transform (ODWT) was then used to compute the wavelet coefficients through iterative time-frequency decomposition. Minutely, the wavelet coefficient's amplitude evaluated the similarity degree between the wavelet and the actual signal, whereas its sign specified the type of the correlation (positive or negative). The discrete version of the wavelet transform was preferred instead of the continuous one, so as to discard redundant information; whereas the orthogonal basis facilitated the perfect reconstruction of the initial brain data. The decomposition scheme, of j = 1…5 levels, engaged recursive low pass filtering to extract the activity of each frequency band with optimal resolution. The above computations were implemented through Matlab functions. The window length was the minimum one that contained at least one coefficient from each frequent band.

For each time window and each electrode, the wavelet coefficients were computed through Eq. (1) based on the decomposition scheme for the five frequency bands. Given that there were multiple coefficients for the decomposition levels (k = 1…K), these coefficients were first squared and then summed to provide the energy of each frequency band (E_j).

$$E_j = \sum_{k=1}^{K} |Ck^2|, j = 1 \dots 5 \qquad (1)$$

where j shows the decomposition level, C_k corresponds to the wavelet coefficient and E_j is the energy value for each brain rhythm. Then, we calculated the total energy (E_{tot}) of the signal by simply summing all the energies for each frequency band:

$$E_{tot} = \sum_{j=1}^{5} E_j \quad (2)$$

In the end, we computed the relative energy for each frequency band by simply dividing its absolute energy value (E_j) with the total EEG energy (E_{tot}) [234]. These relative energies depicted the total energy contribution of a specific rhythmic activity to the energy of the whole EEG. In other words, these numbers present the ratio of each frequency band's energy to the total energy and thus, they are positive and equal to one.

In the majority of cognitive neuroscience studies, researchers use Fast-Fourier Transform (FFT) or morlet wavelets to examine the frequeny components of an EEG signal. However, this work employed the family of bi-orthogonal wavelets of 5th order since this was a suitable choice when dealing with EEG/ERP data for the reasons described above, as also indicated by Frantzidis et al. [235]. Moreover, the ODWT offers excellent time-frequency resolution in comparison with the FFT.

Motion Captures and analysis

For the recording of the motion and the body skeleton, we used Kinect V2 by Microsoft. Since Kinect V2 provides information of x,y, and z positions of 25 joints which are shown in figure 5.5, we can readily detect key body joints for sitting posture (since participants were seated all the time). The analysis of these motion data was conducted mainly by the IMI at the NTU.

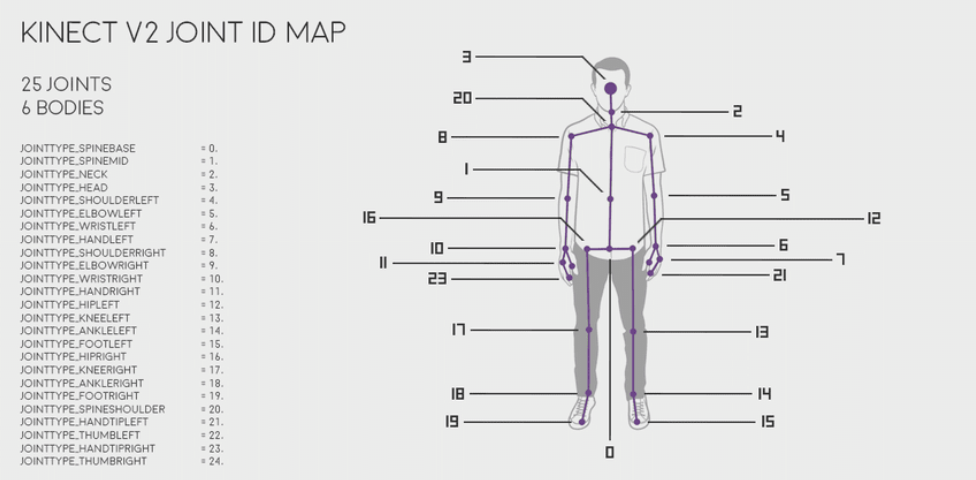


Figure 5.5 The 25 points examined with Kinect

We divided the skeleton into five parts to understand motion in different body part movements as a part of processing shown in Table 5.3.

Motion data was used for two purposes: emotion recognition based on movement and examination of movement changes between the interactions. Emotion recognition will allow us to compare the results with the outcome of the questionnaire. The method proposed by Tomasz Sapiński [237] was used, which is a different representation of affective movements, based on a sequence of joint positions and orientations. The focus was on seven affective states: neutral, sadness, surprise, fear, disgust, anger, and happiness. The algorithm utilizes a sequential model of affective movement based on low-level features, which are positions and orientation of joints within the skeleton provided by Kinect v2. A more detailed description of this methodology can be found in [238].

Table 5.3 The five body parts we examined and the points used for each one

Torso	Right Arm	Left Arm	Right Leg	Left Leg
Neck	ElbowR	ElbowL	KneeR	KneeL
SpineShoulder	WristR	WristL	AnkleR	AnkleL
SpineBase	HandR	HandL	FootR	FootL
ShoulderR				
ShoulderL				
HipR				
HipL				

Secondly, we were interested in the difference of movement among the three cases. Thus, the average difference in degrees of each body part mentioned in Table 4 was calculated. Using information of x,y and z positions of 25 joints for each frame during interactions, we computed the standard deviation along the specified axis for each joint through all the frames with respect to first point,as per equation (3).

For each joint:

{(xi,yi,zi) : i = 1,…n} is a collection of points of movement and (x,y,z) the starting point for given joint, where (x,y,z)= $\frac{1}{n}\sum_{i=1}^{n}(x_i\ ,\ y_i\ ,\ z_i)$. Then,

$$\sigma^2 = \frac{\sum_{i=1}^{n}((x-x_i)^2+(y-y_i)^2+(z-z_i)^2}{n} \quad (3)$$

$\sigma = \sqrt{\sigma^2}$ gives the standard deviation in Euclidean distance of all points from the centroid (starting point).

EMG recordings and analysis

To verify the features extracted from the motion captures, but also to examine the motion from another perspective, we decided to use EMG for selected muscles of the upper body. We used a Myon Aktos wireless EMG system (https://www.myon.ch/aktos) with eight sensors, a sampling rate of 2000Hz, for four preselected muscles (trapezius, biceps, upper abdominal muscles, and abductor of the thumb), as shown in Figure 5.6. After a thorough discussion with a group of doctors from "Arogi Euromedica" Physical Rehabilitation Center in Thessaloniki, Greece we concluded on four muscles that can provide us with substantial information on humans' reaction and behavior during an interaction.

In detail, we start from the trapezius muscle, which is in the upper part of the back and it is responsible for the elevation of the scapula. In other words, this muscle can give us data regarding the lifting of the shoulders that can act as a stress indicator. Secondly, we chose the biceps muscle, which is responsible for the flexion of the arm. As the biceps is one of the main muscles of the arm, we can monitor any possible movement done by the arm. The third muscle is the thenar muscles, which are responsible for the movements of the thumb, like flexion. However, these muscles can reveal any possible small movement done with the hand or fingers. Usually, we move the thumb first when we intend to proceed with a fingers' movement. The fourth and last muscles are the upper abdominal muscles so that we could monitor any flexion or rotation of the trunk, as well as any alteration in the breathing patterns. Moreover, we can monitor if the participant was sitting in a tight, uncomfortable position during any time of the experiment having one more indicator for stress or the sense of not being comfortable. Figure 5.6 shows an example of a participant with the EMG electrodes and the exact position of the muscles.

EMG signal acquisition and processing were conducted through EMG and Motion Tools Software by Cometa [239] with the sampling frequency of 2KHz. So, firstly, we applied a Butterworth high pass filter at 20Hz to exclude the noise caused by the motor units' firing rate [240] and a low pass filter at 500Hz for high. Then, we run a frequency analysis to find the mean frequency (Fmean) for each muscle of each body side. We lastly computed the Root Mean Square (RMS) which quantifies the electric signal as it indicates the physiological activity during contraction [241]. RMS provides information on the muscle activation intensity. RMS and Fmean are the most commonly used variables for the analysis of an EMG signal [241–243].

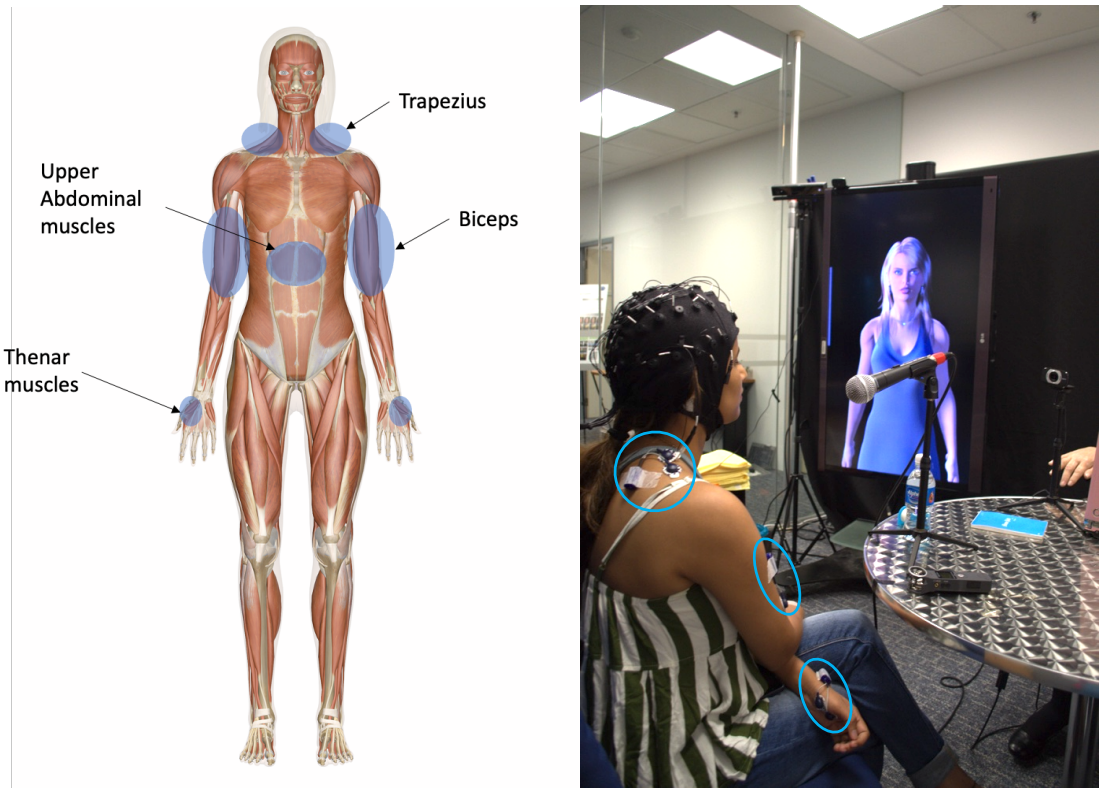


Figure 5.6 LEFT: The four muscles' positions where we put the EMG electrodes. We used 4 electrodes for each side of the body. RIGHT: An example of a participant wearing the physiological equipment (EEG and EMG) while interacting with Nadine. The three positions of EMG electrodes that are obvious are highlighted.

Audio recordings and analysis

For the recording of the sessions, we used an easy-to-use portable recorder called Zoom H1 Handy recorder. The audio was saved in a .wav format, 24-bit, with a sampling rate of 96kHz. The participants were recorded in a sitting position. We recorded both parts of the experiment, which means we have 240 recordings for the 40 participants. However, audio data were mainly analyzed and used for the first part regarding the job interview in the three different types of interaction. The duration of each recording varied for each interaction and each participant from around 2 – 6 minutes.

Given the background of our participants, each style of communication was slightly different. We used Praat software [179] and Matlab. The pre-processing analysis of the signal was conducted in Matlab to clean the possible noise and then, the filtered signal was imported to Praat for further analysis. We annotated the data based on the questions and the thematic areas and we examined each question/answer separately. For each question, we worked with segments of 40.000 samples each. We examined several time-related conversational features and prosodic/acoustic ones, as shown in Table 5.4. The rationale behind the

selection of these voice features was based on already established associations with human emotions, like frequency, timing, and volume [244] and on already efficiently used vocal features [3, 33, 34].

Table 5.4 Acoustic/Prosodic and Conversational Features used. Yellow: time-related features, Grey: Frequency-related features and Volume, Blue: Acoustic features

Acoustic / Prosodic Features			
Human		**Agent**	
Duration of each answer	*T*		
Total duration of each interaction	*TD*		
Pause duration	*PD*		
Fundamental Frequency			*F0*
Minimum Fundamental Frequency			*F0min*
Maximun Fundamental Frequency			*F0max*
Formants			*F1, F2*
Intensity			*I*
Pulses			*Pl*
Jitter			*J*
Shimmer			*S*
Harmonicity (Harmonics-to-noise ratio)			*HNR*

Conversational Features			
Human		**Agent**	
Voice breaks	*VB*		
Response time			*Rt*
Speechrate (nsyll/sec)			*Sr*

In Figure 5.7 we see an example of a female participant sample, interacting with Nadine, as well as a sample of Nadine replying to her. The yellow line represents the intensity in dB, the blue one the fundamental frequency in Hz, and consequently the outcome of the pitch. The red dots are the different formants, based on the F0 and the blow vertical lines illustrate the pulses.

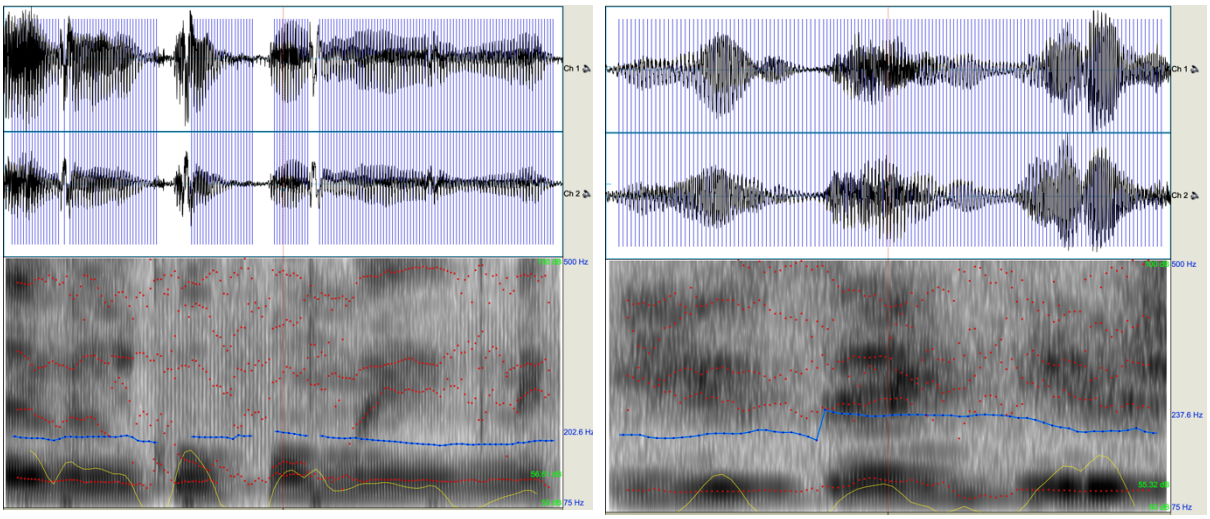


Figure 5.7 LEFT: Audio sample of a female participant interacting with the social robot Nadine in Praat software. RIGHT: Sample of Nadine's voice while replying to the participant's answer. Up: the voice signal with the voice breaks. Down: the spectrogram where the yellow line indicates the intensity of the voice in dB, the blue line the fundamental frequency in Hz, and the red line-dots the several formants.

Emotion recognition through audio

To complement our research on humans' emotional states, we also conducted an emotion recognition based on the audio signal. Inspired by the work of Trigeorgis et al. [245], we built a Machine Learning model that could detect emotions from the participants' voices. Same voice segments as before were used. For our training purposes, we used an already trained dataset, called RAVDESS [246], which includes 1500 audio files from 24 different actors, 12 males and 12 females. Convolution Neural Network, which has been tackled by the use of Long Short-Term Memory (LSTM), was used for classification purposes. Lastly, for feature extraction python library LibROSA [247] was used. The model detected emotions with more than 70% accuracy. We examined 8 different emotions i.e boredom, calm, happiness, anger, fear, disgust and surprise.

ML Classification analysis

To complement the above and to highlight the differences in the vocal behavior, we developed an ML model that can recognize the nature of the interlocutor a human speaks with, taking as input the human vocal behavior. We kept the separation of our data in three classes, according to the nature of the interlocutor. Our goal is to separate the three classes and predict the class of new samples. However, our dataset was imbalanced, due to the number of questions asked in each interaction.

Participants' reactions to the job interview questions were recorded and labeled according to the nature of the interlocutor. The audio recordings were divided into shorterm windows of 100ms and for each, a total

of 34 distinct features were extracted. Additional statistical features, like standard deviation and average values, were added to calculate changes over time or other differences and thus in total, we have 136 distinct features per sample. To facilitate our work, we used the "pyAudioAnalysis" library in python to extract the audio features [248]. Audio acoustic features both from time and frequency domain were used. Having verified that all features follow a Gaussian distribution, data were standardized.

Subsequently, multiple dimensionality reduction methods were compared to find the one that suits best our needs. We tested 5 distinct methods, namely Principal Component Analysis (PCA), Singular Value Decomposition (SVD), Linear Discriminant Analysis (LDA), Isomap Embedding (ISO), and Locally Linear Embedding (LLE). After each reduction, a Support Vector Machine Classifier (SVM) with Radial Basis Function (RBF) kernel was fitted to the data. Results were acquired via the Stratified 10-Fold Cross Validation (CV) to compare the dimensionality reduction methods. PCA, SVD, and LDA performed significantly better than the other methods, as shown in Table 5.5. Finally, LDA was selected since it outperformed all other methods.

Table 5.5. Scores of dimensionality reduction methods

n_components	20	40	60	80	100
PCA	0.531	0.576	0.618	0.633	0.640
SVD	0.538	0.569	0.624	0.637	0.637
LDA (n=2)	**0.803**	**0.803**	**0.803**	**0.803**	**0.803**

Note: Score is F1-macro of the test dataset

Having assured the performance of the dimensionality reduction, we tested different models to find the one that will better fit to our data and will perform the best classification. We tested several models as shown in table 5.6. A 10-Fold Stratified Cross Validation (CV) was employed for the comparison of these models. The K-Nearest Neighbors (KNN) Classifier outperformed all other classifiers tested. KNN represents the numbers of neighbors/samples (K) that participate in the voting of classifying a new point. To fine tune this parameter, different values were tested from the range of 5-60 and we found that K=15 is the optimal value. Finally, a comparison of the three dimensionality methods of table 5.5 was repeated using the selected classifier, which verified that LDA is the best method for our task. To wit, results for PCA and SVD flunctuated around 0.60 (60%) F1-score compared to 0.82 (82%) F1-score of LDA.

As a model metric, accuracy is one of the most interpretable, straightforward metrics, but it is dangerously misleading when used on an imbalanced dataset. Thus, the macro F1-score was used to compare and evaluate the different models. F1 score is a function of Precision and Recall. Precision is the ratio of correctly classified samples per class to the total classified samples per class [249]. Thus a value of precision is calculated for each class. Then weighted precision is calculated, which is the average precision between

classes that takes class imbalance into account. Recall is the ratio of correctly classified samples per class to the total number of samples per class [249]. As before, only weighted recall is shown in this work. Thus, the F1 score is the harmonic mean of precision and recall. Weighted F1 score is the weighted average F1 score of each class. In our case, the F1 score is used to show the results instead of accuracy since our dataset is imbalanced, along with precision and recall.

Table 5.6 Comparison of the different models

	Accuracy	**Precision weighted**	**Recall weighted**	**F1 score weighted**
SVC linear	0.801	0.806	0.801	0.8
SVC poly	0.777	0.793	0.777	0.773
SVC sigmoid	0.718	0.726	0.718	0.718
SVC rbf	0.808	0.813	0.808	0.807
Decision Tree Classifier (maximum depth = 2)	0.805	0.825	0.805	0.807
AdaBoost (with DTC and maximum depth = 10)	0.784	0.789	0.784	0.783
MLP Classifier (single layer of 10 neurons)	0.806	0.812	0.806	0.805
Gaussian Naive Bayes Classifier	0.808	0.813	0.808	0.806
Nearest Neighbor Classifier (k = 15)	0.820	0.827	0.820	0.820

Psychometric data

To complement the collected physiological data, we used also two types of questionnaires to assess the human attitude towards robots and emotions as well as perception during the interactions.

The first questionnaire is the valid Negative Attitudes Towards Robots scale (NARS), firstly introduced by Nomura et al [218]. This questionnaire, or its subscales, have already been used successfully by several studies to assess humans' attitude towards robots. Participants had to reply to its questions before the onset of the experiment, thus before their interaction with Nadine.

In the meantime we used a reliable, valid questionnaire, including closed-ended Likert-scale questions, to assess humans' emotions, mood states, and the overall experience of the interactions. The emotion scale was based on the Positive and Negative Affect Schedule with two scales: one measuring positive affect and

the other measuring negative. The emotions were based on the emotions scale PANAS X [250]. The questionnaire was given to the participants before as well as after the experiment so that we could define any possible differences in their state or their expectations.

Statistical analysis

The statistical analysis for all modalities was conducted in SPSS. We controlled the internal consistency, and we assured the normality of our data (Kolmogorov-Smirnov and Shapiro-Wilk tests), which led us to a parametric repeated measures ANOVA. However, some modalities suggested us a different type of ANOVA as between-group variables, like gender and ethnicity, had an interaction with our data. In such cases, we used a mixed ANOVA and we controlled also the equality of variances through Levene's test. A Greenhouse-Geisser correction was applied when violations of the sphericity assumption were noticed. The statistically significant results of the ANOVAs were followed up by the Bonferroni post hoc tests.

Pearson correlation and simple linear regression were also used to examine correlations and linear relationships among our data.

5.2.4 Human reactions and behaviors during H-H and H-NH Interactions

The first part of our experiment aims to use human's complete behavior to reveal the humans' needs towards human-nonhuman communication and to find emotional and behavioral patterns that can enhance the field of HRI. Our results will also allow us to examine the effects of a robot-mediated job interview, concluding with a bifold assessment: if nonhuman agents are really in the position of offering more meritocratic interviews and if humans would indeed prefer such a case.

Voice and body reactions

Brain activity measured by EEG

As described in the chapter of Methodology, we calculated the relative energy for each frequency band through ODWT. Results are shown in Table 5.7 and Figure 5.8. Table 1 depicts the mean values along with the standard deviations (SD) and the statistically significant differences between the interactions. Minutely, we can see that in Prefrontal cortex no difference is significant between the three conditions: : $F(2, 64) = 0.282$, $p = 0.755$, $\eta^2 = 0.01$ for delta, $F(2, 64) = 1.949$, $p = 0.151$, $\eta^2 = 0.057$ for theta, $F(2, 64) = 0.625$, $p = 0.539$, $\eta^2 = 0.019$ for alpha, $F(2, 64) = 1.027$, $p = 0.364$, $\eta^2 = 0.031$ for beta, and $F(1.421, 45.459) = 2.096$, $p = 0.131$, $\eta^2 = 0.061$ for gamma. For the last comparison, the Greenhouse-Geisser correction was

applied due to a violation of the sphericity hypothesis. However, we can see that delta and theta bands are the highest in all conditions.

Table 5.7 Differences in the brain states of each brain area among interactions

Brain areas	Brain states	H-H		H-NA		H-VH	
		M(SD)		M(SD)		M(SD)	
Prefrontal	delta	0.225	(0.052)	0.231	(0.062)	0.232	(0.046)
	theta	0.282	(0.036)	0.285	(0.032)	0.296	(0.030)
	alpha	0.182	(0.028)	0.175	(0.026)	0.179	(0.021)
	beta	0.177	(0.026)	0.181	(0.034)	0.172	(0.026)
	gamma	0.123	(0.013)	0.128	(0.028)	0.119	(0.016)
Frontal	delta	0.211	(0.045)	0.189	(0.047)	0.215	(0.047)
	theta	0.216	$(0.034)^{c}$	0.270	$(0.030)^{a,b}$	0.297	$(0.038)^{b}$
	alpha	0.234	$(0.029)^{c}$	0.176	$(0.021)^{a}$	0.181	$(0.020)^{a,b}$
	beta	0.204	$(0.026)^{c,b}$	0.211	$(0.032)^{a,c}$	0.187	$(0.031)^{b}$
	gamma	0.125	$(0.016)^{c}$	0.155	$(0.024)^{a}$	0.120	$(0.024)^{c,b}$
CentroParietal	delta	0.206	(0.051)	0.219	(0.055)	0.217	(0.045)
	theta	0.277	(0.038)	0.291	(0.036)	0.293	(0.035)
	alpha	0.195	(0.033)	0.186	(0.025)	0.187	(0.025)
	beta	0.186	(0.029)	0.184	(0.034)	0.181	(0.028)
	gamma	0.122	(0.019)	0.121	(0.024)	0.120	(0.026)
Parietal	delta	0.215	(0.048)	0.215	(0.041)	0.214	(0.045)
	theta	0.286	$(0.032)^{c}$	0.186	$(0.032)^{b,a}$	0.193	$(0.038)^{a}$
	alpha	0.198	$(0.031)^{c}$	0.290	$(0.027)^{b,a}$	0.302	$(0.023)^{a}$
	beta	0.181	(0.021)	0.182	(0.029)	0.176	(0.023)
	gamma	0.114	$(0.020)^{c}$	0.127	$(0.027)^{b,a}$	0.115	$(0.022)^{a}$
Temporal	delta	0.192	(0.041)	0.198	(0.048)	0.197	(0.051)
	theta	0.250	$(0.033)^{c}$	0.200	$(0.030)^{b,a}$	0.204	$(0.032)^{a}$
	alpha	0.199	$(0.028)^{c}$	0.285	$(0.021)^{b,a}$	0.294	$(0.022)^{a}$
	beta	0.234	$(0.021)^{c}$	0.187	$(0.033)^{b,a}$	0.181	$(0.029)^{a}$
	gamma	0.125	(0.015)	0.130	(0.021)	0.123	(0.023)
Occipital	delta	0.194	$(0.046)^{c}$	0.219	$(0.050)^{b,a}$	0.220	$(0.051)^{a}$
	theta	0.224	$(0.045)^{c}$	0.183	$(0.033)^{b,a}$	0.188	$(0.044)^{a}$
	alpha	0.167	$(0.037)^{c}$	0.289	$(0.025)^{b,a}$	0.302	$(0.023)^{a}$
	beta	0.262	$(0.042)^{c}$	0.181	$(0.022)^{b,a}$	0.170	$(0.021)^{a}$
	gamma	0.160	$(0.029)^{c}$	0.128	$(0.030)^{b,a}$	0.119	$(0.025)^{a}$

Note: Means in the same row with different letters are significantly different from one another based on post hoc t-tests with the Bonferroni correction ($p<0.05$). The absence of letters signifies no statistical significance.

In the Frontal cortex, we found a statistically significant difference in both theta and alpha band for the H-H and H-nonhuman (NH) interaction. To wit, theta state is dominant in H-NA and H-VH interaction with a significant difference from H-H whereas alpha state is dominant in H-H with a significant difference from the other two conditions ($F (2, 64) =5.597$, $p =0.006$, $\eta^2 = 0.149$ and ($F (2,64) = 17.147$, $p <0.001$, $\eta^2 = 0.461$ respectively). Gamma band is also significant higher during H-NA interactions compared to H-H and H-VH ($F (1.46,46.76) =12.925$, $p<0.001$, $\eta^2 = 0.288$) and delta band during H-NA and H-H ($F (2, 64) =2.769$, $p =0.003$, $\eta^2 = 0.144$ and ($F (2,64) = 3.121$, $p=0.004$, $\eta^2 =0.120$, respectively).

For the CentroParietal cortex, there were no significant differences between the conditions in any of the brain state. $F (2, 64) = 0.980$, $p = 0.381$, $\eta^2 = 0.030$ for delta, $F (2, 64) = 2.146$, $p = 0.125$, $\eta^2 = 0.063$ for theta, $F (2, 64) = 0.939$, $p = 0.396$, $\eta^2 = 0.029$ for alpha, $F (2, 64) = 0.330$, $p = 0.720$, $\eta^2 = 0.010$ for beta, and $F (1.667, 53.347) = 0.207$, $p = 0.814$, $\eta^2 = 0.006$ for gamma. For this last comparison, the Greenhouse-Geisser correction was applied due to a violation of the sphericity hypothesis.

Regarding the Parietal cortex, alpha band found to be significant higher in both nonhuman interactions with a significant difference compared to H-H interaction ($p<0.001$) where theta state has the dominant role ($p<0.001$) ($F (2,64) = 28.209$, $p<0.001$, $\eta^2 =0.548$ and $F (2,64) = 18.632$, $p<0.001$, $\eta^2 = 0.446$ respectively). Gamma state ($F (2, 64) = 3.493$, $p = 0.036$, $\eta^2 = 0.098$) was also higher during H-NA interaction, significant different from the two other conditions ($p =0.005$). ANOVAs exhibited no significant results for delta ($F (2, 64) = 0.003$, $p = 0.997$, $\eta^2 = 0.000$) and beta ($F (2, 64) = 1.094$, $p = 0.341$, $\eta^2 = 0.033$) frequency bands.

In the Temporal cortex, theta band found to be significant higher during H-H interaction compared to H-NA and H-VH ($p<0.001$) where alpha band is the one significant higher ($F (2, 64) = 5.493$, $p = 0.003$, $\eta^2 = 0.298$ and $F (2,64) = 26.129$, $p<0.001$, $\eta^2 =0.601$ respectively). The high value of η^2 confirms the significance. The beta band also presented a significant difference ($F (2,64) = 22.131$, $p<0.001$, $\eta^2 =0.677$) during H-H with significance from H-NA ($p=0.002$) and H-VH ($p<0.001$) with also a high value of η^2.

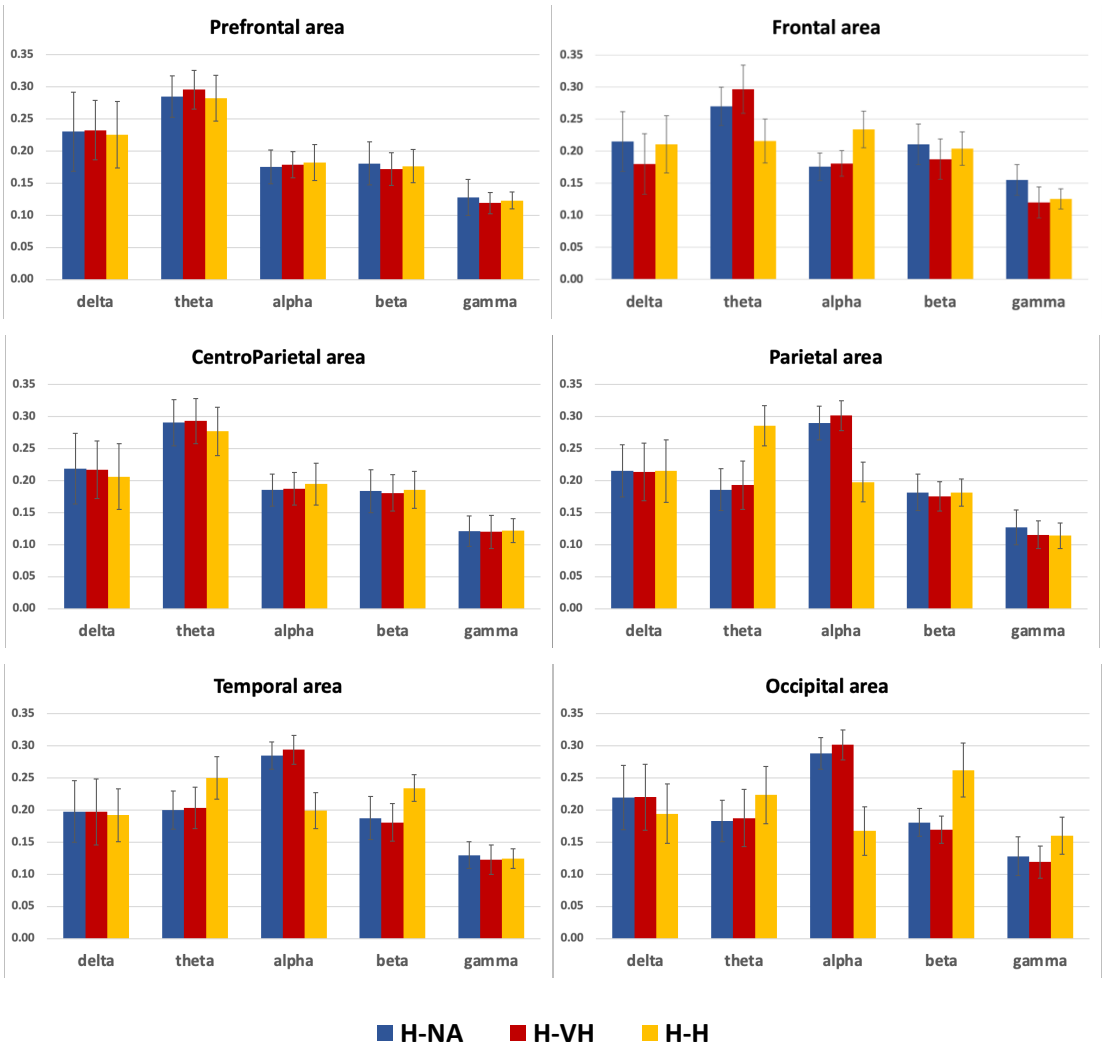


Figure 5.8 Relative energies for the five brain states in the six brain areas for the three interactions

Lastly, in the occipital cortex, the alpha band, higher in both H-NA and H-VH, presented a highly significant difference with the H-H ($p<0.001$), where the beta state found to be the dominant one with a high significant difference from the other two conditions as well ($p<0.001$) ($F(1.59, 50.887) = 157.755$, $p<0.001$, $\eta^2= 0.831$ and $F(2,64) = 202.523$, $p < 0.001$, $\eta^2 = 0.864$ respectively). Theta state found also to be higher in H-H interaction ($F(2, 64) = 19.059$, $p <0.001$, $\eta^2= 0.373$), significantly different from H-NA and H-VH ($p<0.001$). A small difference was also found in the delta band ($F(2, 64) = 4.635$, $p = 0.013$, $\eta^2= 0.109$). The histograms of Figure 5.8 clearly illustrate these results with the relevant SDs.

Table 5.8 Brain states per interaction

	H-H M(SD)	H-NA M(SD)	H-VH M(SD)
delta	0.211 (0.043)^{a,c,d}	0.192 (0.035)^{a,d}	0.196 (0.037)^{a,d}
theta	0.268 (0.042)^{b}	0.236 (0.035)^{b,c}	0.245 (0.023)^{b,c}
alpha	0.189 (0.021)^{c,a,d}	0.263 (0.048)^{c,b}	0.261 (0.034)^{c,b}
beta	0.193 (0.025)^{d,a,c}	0.178 (0.026)^{d,a}	0.178 (0.023)^{d,a}
gamma	0.129 (0.019)^{e}	0.132 (0.022)^{e}	0.119 (0.021)^{e}

Note: Means in the same column with different letters are significantly different from one another based on post hoc t-tests with the Bonferroni correction ($p<0.05$).

We also calculated the mean relative energy in each interaction, shown in Table 5.8. For the H-H interaction, the repeated measures ANOVA showed a significant difference between the brain states ($F(2.606, 88.610) = 70.944$, $p<0.001$, $\eta^2= 0.676$), with the dominant state to be the theta one (Er = 0.268 ± 0.042) and the second higher the beta. During H-NA interaction, we also found a significant difference in the mean energy of the brain states ($F(2.845,93.873) = 84.203$, $p<0.001$, $\eta^2 = 0.718$) with the alpha band to be the dominant and theta the second one, with a significant difference between them. The same results were found for the H-VH interaction. In both H-NH interactions, beta and gamma states were significantly lower than the others with no difference between them. In general, we can see that for both nonhuman interactions, the alpha state prevails whereas during H-H interaction we noticed mainly the theta rhythm. Figure 5.9 illustrates the differences in brain states per interaction.

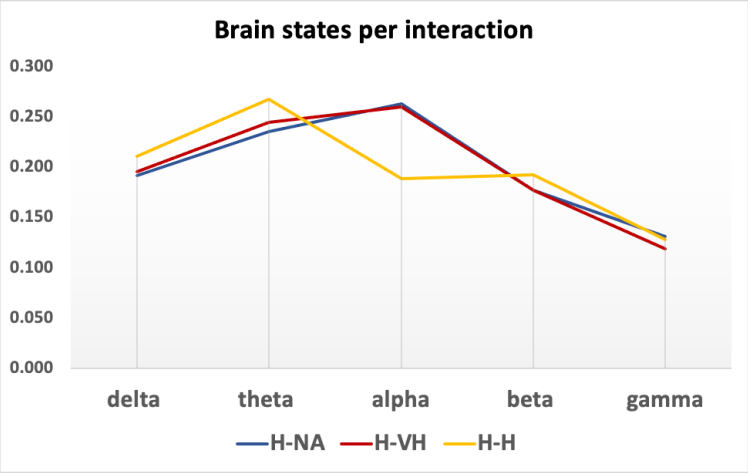


Figure 5.9 Differences in brain states per interaction.

Motion data

Regarding our motion data, extracted from the Kinect, we firstly examined how our body movements change based on the nature of our interlocutor. Thus, we calculated the degree of movement's range for each joint to investigate how much each predefined body point was moved, as shown in Figure 5.10. To wit, movement's range shows us how each body joint has moved during the interaction or, in other words, how much its starting point has moved.

Movement does not change in the same way for each interaction. We can see that during H-H, the upper body follows smoother changes that present no significant differences between them. Consequently, there is no difference between the two sides, neither between the upper and the lower part except the hip movement but there is a significant difference among all the body points ($F(3.78, 139.89) = 12.070$, $p < 0.001$, $\eta^2 = 0.246$). Regarding H-NA and H-VH interactions, there is also a significant difference among all body points with $F(4.33, 160) = 25.153$, $p < 0.001$, $\eta^2 = 0.405$, and $F(3.21, 118.7) = 27.996$, $p < 0.001$, $\eta^2 = 0.431$ respectively. We can see that the left side of the upper body shows a significantly bigger range of movement for both interactions. Same for the lower body compared to the upper one.

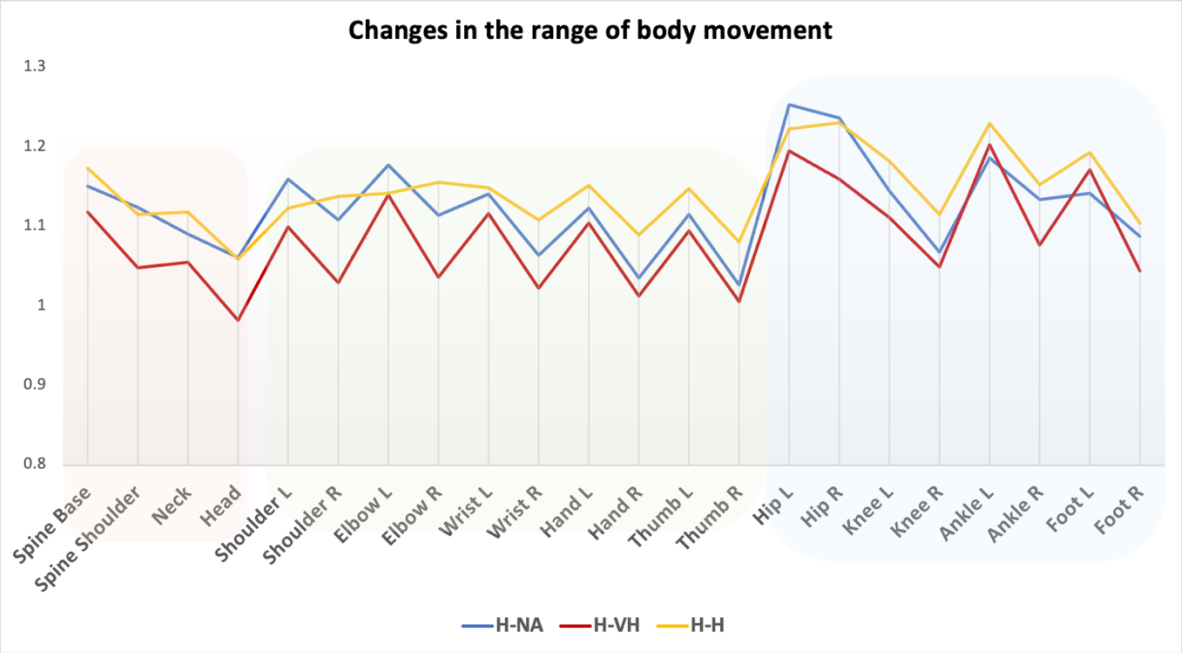


Figure 5.10 Changes in movements' range during the three interactions. The three different colored shades divide the three basic body structures (red – body trunk, green – upper body, blue – lower body)

From the comparison of each body part among the three interactions, we found that there is no statistical significance in the left side of the human body among the interactions. However, for the right side of all the

body parts, the difference is significant only between H-H and H-VH, except the parts of the shoulder and elbow where there is also a difference between H-H and H-NA.

Table 5.9 presents in detail the differences and the statistical significances in the range of movement between the three cases.

Table 5.9 Differences in the range of movement between the conditions H-NA, H-VH and HH

	H-NA M (SD)		H-VH M (SD)		H-H M (SD)	
Spine Base	1.151	(0.138)[a,b]	1.118	(0.082)[a]	1.173	(0.123)[b]
Spine Shoulder	1.124	(0.117)[b,c]	1.048	(0.073)[a]	1.115	(0.116)[b]
Neck	1.090	(0.128)[a,b]	1.055	(0.072)[a]	1.118	(0.114)[b]
Head	1.061	(0.129)[b,c]	0.982	(0.087)[a]	1.058	(0.121)[b]
Shoulder L	1.159	(0.111)	1.099	(0.084)	1.123	(0.118)
Shoulder R	1.108	(0.114)[a]	1.029	(0.063)[a]	1.137	(0.114)[b]
Elbow L	1.177	(0.117)	1.139	(0.090)	1.142	(0.115)
Elbow R	1.114	(0.124)[b,c]	1.036	(0.057)[a]	1.155	(0.117)[b]
Wrist L	1.141	(0.131)	1.116	(0.081)	1.149	(0.126)
Wrist R	1.064	(0.139)[a,b]	1.022	(0.084)[a]	1.108	(0.125)[b]
Hand L	1.123	(0.142)	1.105	(0.082)	1.151	(0.142)
Hand R	1.035	(0.155)[a,b]	1.012	(0.108)[a]	1.090	(0.136)[b]
Thumb L	1.115	(0.142)	1.094	(0.083)	1.148	(0.143)
Thumb R	1.027	(0.157)[a,b]	1.006	(0.109)[a]	1.080	(0.138)[b]
Hip L	1.253	(0.125)[a,b]	1.195	(0.080)[a]	1.223	(0.117)[b]
Hip R	1.236	(0.124)[b,c]	1.160	(0.084)[a]	1.231	(0.112)[b]
Knee L	1.145	(0.169)	1.111	(0.134)	1.182	(0.169)
Knee R	1.068	(0.162)[a,b]	1.049	(0.091)[a]	1.115	(0.160)[b]
Ankle L	1.186	(0.189)	1.202	(0.129)	1.230	(0.192)
Ankle R	1.133	(0.165)[a,b]	1.077	(0.140)[a]	1.152	(0.161)[b]
Foot L	1.142	(0.204)	1.171	(0.143)	1.193	(0.202)
Foot R	1.087	(0.174)	1.044	(0.160)	1.104	(0.167)

Note: Means in the same row with different letters are significantly different from one another based on post hoc t-tests with the Bonferroni correction ($p<0.05$). The absence of letters signifies no statistical significance. Values are measured in degrees.Emotion recognition

Except the range of movement, we also conducted an emotion recognition, which gave us the percentages of the seven discreet emotions that can verify or contradict the results of the questionnaire. Statistical significance is shown in the results of Figure 5.11 with the use of the asterisks.

Emotions during H-H interaction have always a statistical significance compared to the other two conditions. Happiness and fear are the two dominant emotions during H-H but also surprise presents a higher level. H-NA interaction is described mainly by a neutral emotional situation but also by fear and anger. Fear has no statistically significant difference between H-H and H-NA conditions. H-VH interaction presents an equal neutral emotional state to H-NA, but we also found the lowest level of positive emotions.

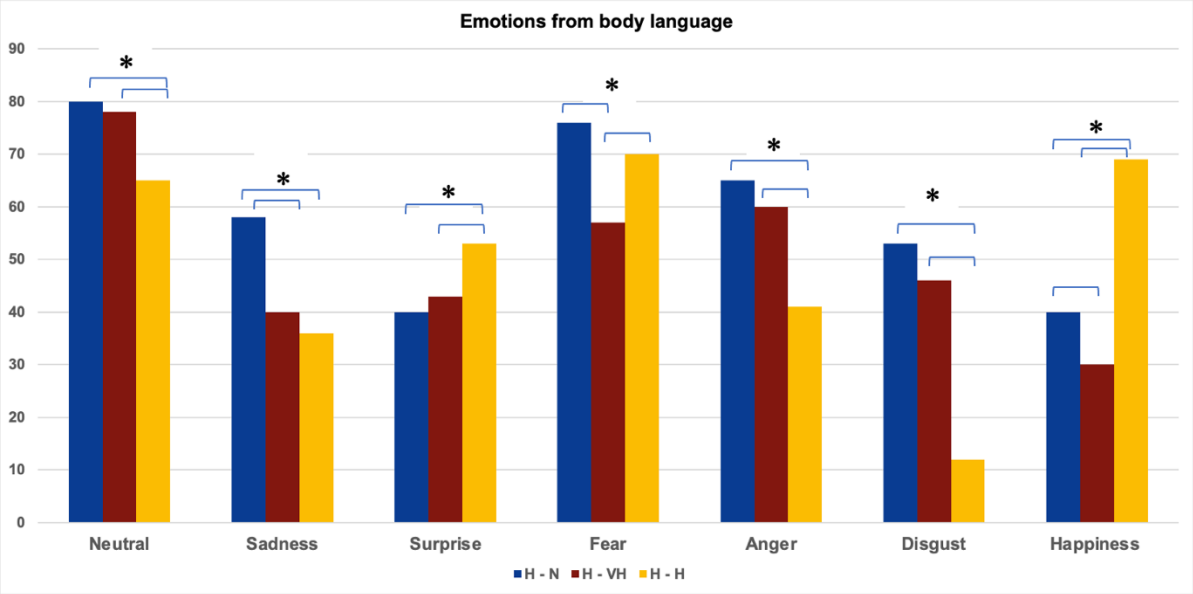


Figure 5.11 Percentages of emotions extracted from the body movement captured by Kinect in the three interactions.

Muscles' activity measured by EMG

We first conducted a frequency analysis to calculate the mean value of each of the four predefined muscles, for both sides. Results are depicted in Table 5.10, along with their statistical significance. We noticed a higher activity of the muscles in the right side, especially during the interaction with the nonhuman agents. During H-H interaction, our repeated measures ANOVA exhibited a significant difference among all muscles ($F(4.5, 170.8) = 27.745$, $p<0.001$, $\eta^2=0.413$) but only the muscle of the shoulder (trapezoid) had a difference between the two sides ($p=0.003$). During H-NA and H-VH interactions, the difference was significant among all muscles except from the thenar one ($F(5.19, 192.13) = 33.432$, $p < 0.001$, $\eta^2= 0.575$ and $F(5.16, 185.81) = 35.522$, $p< 0.001$, $\eta^2= 0.588$, respectively) but all of them presented a difference between the two body sides. We also noticed that the difference between the two body sides is smoother during the H-H interaction.

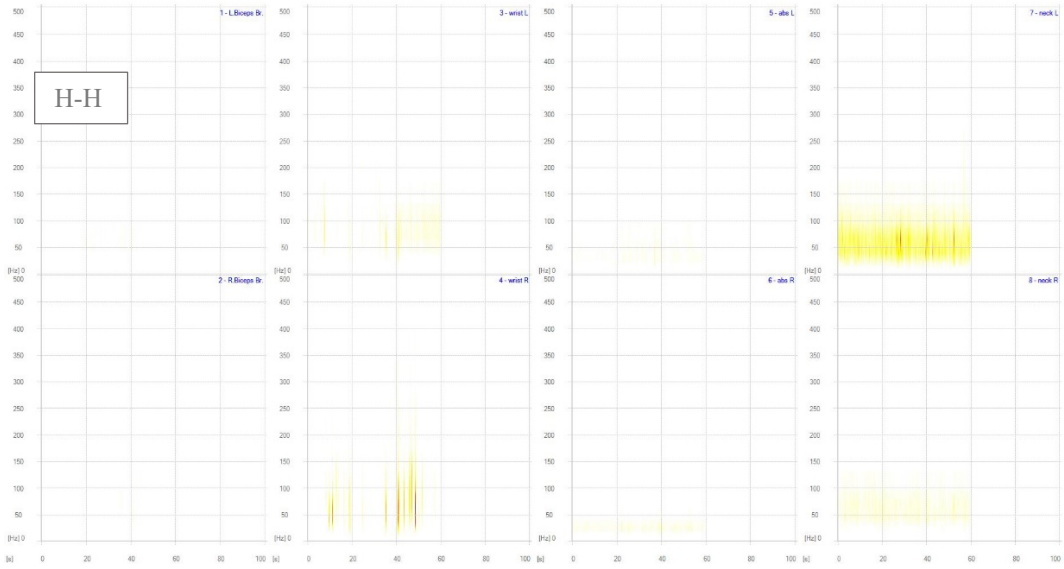
Moreover, we examined possible differences in each muscle between the interactions. We found differences mainly between H-H and H-NA, and mainly for the right side. No differences were found between H-NA and H-VH.

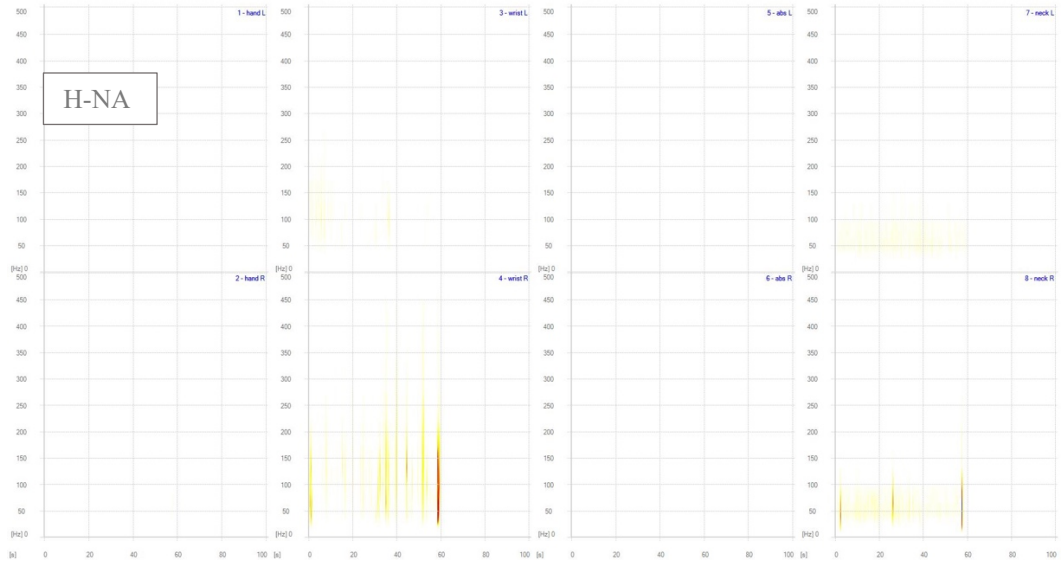
The activation of muscles can be also shown in Figure 5.12 which illustrates an example of a spectrogram extracted from the wavelet analysis. We can see the firing of the muscles for both body sides in each interaction. The spectrogram confirms the results of Table 5.10 with the higher firing in the muscles of H-NH interactions. Specifically, we can see the bigger activation of the trapezius muscle in H-VH, thenar muscles in H-NA and H-VH, and abdominal in H-VH. We can also notice the slightly higher firing of the Biceps muscle in H-VH interaction.

Table 5.10 Mean Values (SD) for the mean frequency (Fmean) and the root mean square (RMS) for both sides of the four muscles per interaction. Bic: Biceps, Tra: Trapezius, The: Thenar and Abs: Abdominal

	Fmean (Hz) M (SD)						RMS (uV) M (SD)					
	H-NA		**H-VH**		**H-H**		**H-NA**		**H-VH**		**H-H**	
BicL	100.3	(19.7)	100.3	(21.6)	95.7	(12.1)	19.2	(3.2)^{a}	17.1	(2.1)^{a,b}	14.4	(3.1)^{b}
BicR	132.6	(22.9)^{a}	125.0	(13.3)^{a,b}	92.8	(11.3)^{c}	24.4	(4.2)^{a}	23.3	(3.5)^{a,b}	21.2	(3.8)^{b}
TraL	30.7	(9.2)^{a}	30.5	(8.2)^{a,b}	18.6	(2.3)^{b}	24.4	(5.0)^{a}	22.5	(7.7)^{a,b}	19.0	(7.1)^{b}
TraR	72.0	(3.6)^{a}	78.3	(7.3)^{a,b}	55.8	(19.7)^{c}	35.3	(4.6)^{a}	33.3	(5.2)^{a,b}	29.0	(4.1)^{b}
TheL	59.8	(10.1)^{a}	53.3	(3.6)^{a,b}	41.5	(5.1)^{c}	32.8	(11.1)	32.5	(11.3)	30.7	(9.2)
TheR	60.0	(9.1)^{a}	52.5	(1.4)^{a,b}	45.5	(6.7)^{b}	53.0	(8.9)^{a}	51.3	(9.2)^{a,b}	42.5	(9.6)^{b}
AbsL	12.8	(3.2)	14.8	(4.6)	11.6	(2.1)	16.7	(3.6)^{a}	15.3	(3.9)^{a,b}	14.1	(4.3)^{b}
AbsR	35.7	(6.3)^{a}	39.7	(10.3)^{a,b}	12.1	(4.1)^{b}	15.5	(3.5)	14.6	(4.3)	14.3	(5.6)

Note: Means in the same row with different letters are significantly different from one another based on post hoc t-tests with the Bonferroni correction ($p<0.05$). The absence of letters signifies no statistical significance. Values are measured in Hz.


H-H
1 - L Biceps Br.
2 - R Biceps Br.
3 - wrist L
4 - wrist R
5 - abs L
6 - abs R
7 - neck L
8 - neck R


H-NA
1 - hand L
2 - hand R
3 - wrist L
4 - wrist R
5 - abs L
6 - abs R
7 - neck L
8 - neck R

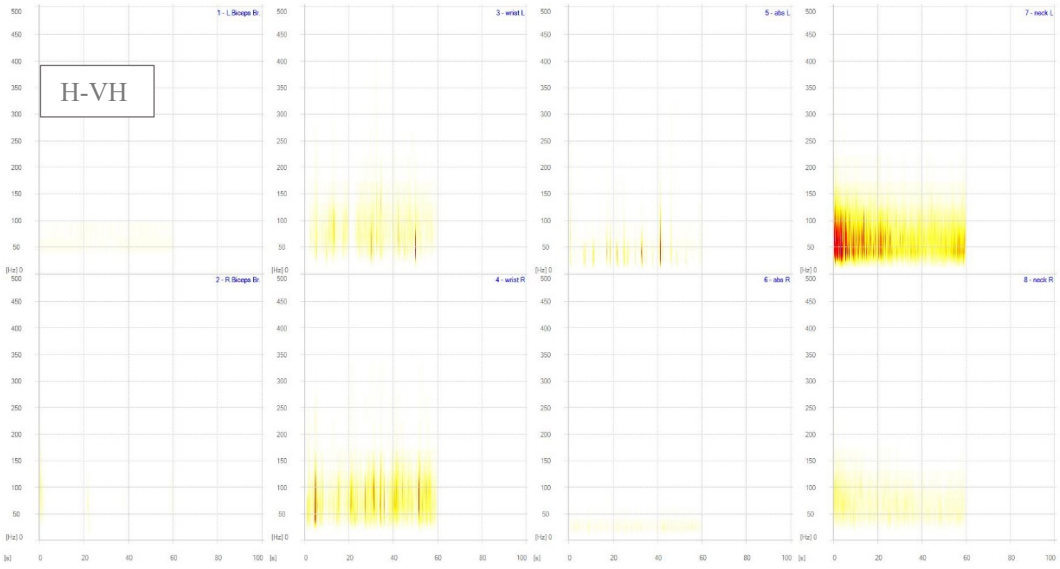


Figure 5.12 Example of a participant's spectrogram extracted from the wavelet analysis for the four muscles on each side and each interaction.

Lastly, we calculated the value of root mean square (RMS), which gives us information about the muscles' intensity, as also shown in Table 5.10. We noticed again a higher intensity in the right side for all muscles and all interactions, except the abdominal where no significant difference was found. Specifically, in H-H, a significant difference among all muscles ($F(2.583, 95.555) = 35.144$, $p<.001$, $\eta^2=0.487$) was found and for both sides ($p<0.001$ for all muscles), except the abdominals. The same result we received for both H-NA and H-VH ($F(2.251, 83.282) = 58.520$, $p<0.001$, $\eta^2=0.613$ and $F(3.648, 134{,}979) = 95.573$, $p<0.001$, $\eta^2=0.721$ respectively). For all comparisons, a Greenhouse-Geisser correction was applied due to violations of the sphericity assumption. A higher intensity was found for the muscle of thenar for all interactions, whereas lower intensity for the abdominals, with similar values among interactions. Differences were significant mainly between H-H and H-NA, as for Fmean.

Audio signal

From the audio signal, we extracted features as shown in Table 3 above, separated in time – related features, frequency – related features where we include the Intensity, and acoustic ones. We examined also each question separately to evaluate their selection and the overall job interview process. Moreover, wherever possible, we explored the reactions of the nonhuman agents, to facilitate and validate the comparison between the interactions.

- Time – related features

Time – related prosodic and conversational features include the duration of each answer, the total duration of each interaction, the pause duration, voice breaks, response time and speechrate. Table 5.11 depicts the results for each interaction with their descriptive statistics.

Specifically, firstly we examined the duration of each response. That means strictly the time starting when participants began to reply to the question until their last word. Expressions of hesitation or uncertainty at the beginning of a response were taken into account. The longest average time for an answer was noted during H-H interaction with a significant difference among the three interactions ($F(2,54) = 44.100$, $p = <0.001$, $\eta^2 = 0.620$). Post hoc comparisons indicated that there is a significant difference both between H-H and H-N and, H-H and H-VH.

The total duration of each interaction includes the welcome and the goodbye of the agent. H-VH interaction presented the longest total duration of all interactions and was found significantly different from H-H and H-N with $p=0.018$ and $p <0.001$ respectively. In general, a significant difference among the three interactions was found ($F(2, 54) = 7.805$, $p=0.001$, $\eta^2 = 0.224$).

Pause duration refers to the average value of all the pauses done during the speech of the participant. We noticed the lower average value during H-H interaction. A significant difference for the three interactions was found ($F(2,54) = 4.699$, $p = 0.013$, $\eta^2 =0.148$) but the post hoc comparisons showed small differences between the pair of groups for the three cases.

In Praat software, voice breaks are described as "the number of distances between consecutive pulses" [179]. H-H presents the highest value and there is a significant difference among the three interactions, ($F(1.3, 35.9) = 31.863$, $p < 0.001$, $\eta^2 = 0.541$). Post hoc comparisons verified the high significance.

Response time refers to the time participants needed to answer a question. Specifically, it starts directly after the end of the agent's sentence until the first sign of response. The lower value was found for H-H whereas in H-N and H-VH the value was significantly higher ($F(2,54) = 49.411$, $p <0.001$, $\eta^2 = .0662$). Pairwise analysis showed a difference between H-H and H-N and H-H and H-VH ($p <.001$ for both). This feature was also examined in the speech of nonhuman agents. The comparison between humans' responses in H-H and the responses of the agents (Nadine, Nicole) showed a significant difference with $F(2,54) = 53.828$ and $p <0.001$.

Lastly, we examined the speech rate, as the number of syllables per second. The feature presented a significant difference among the interactions ($F(2,54) = 11.230$, $p <.0001$, $\eta^2 = 0.294$) and the post-hoc comparisons specified them between H-H and H-N as well as between H-N and H-VH. As before, we conducted the comparison between humans' and agents' responses and we found a significant difference among the interactions, $F(2,54) = 10.422$, $p <0.001$ and $\eta^2 = 0.278$. Post hoc comparisons showed that VH

showed significant differences with both other interactions which means that Nicole had a faster speech rate compared to human participants, as well as to Nadine.

Table 5.11 Summary of the descriptive statistics for the significant acoustic/prosodic and conversational time-related features for the three interactions and the comparison of human/agents

Features	Interactions	Mean	SD	p		Agents	Mean	SD	p
Time (sec)	H-H	15.754	4.159	<0.001					
	H-N	10.464	4.501						
	H-VH	10.396	4.046						
Total Duration (min)	H-H	3.563	0.920	=0.001					
	H-N	3.347	1.121						
	H-VH	4.006	1.081						
Pause Duration (sec)	H-H	0.229	0.059	= 0.013					
	H-N	0.272	0.116						
	H-VH	0.279	0.124						
Voice Breaks	H-H	30	8	<0.001					
	H-N	20	9						
	H-VH	19	8						
Response Time (sec)	H-H	1.147	0.353	<0.001					<0.001
	H-N	2.018	0.545			NA	1.960	0.599	
	H-VH	2.082	0.765			VH	2.174	0.388	
Speechrate (nsyll/sec)	H-H	3.570	0.352	<0.001					<0.001
	H-N	3.260	0.412			NA	3.737	0.651	
	H-VH	3.467	0.507			VH	4.474	0.840	

- Frequency – related features and Intensity

As frequency-related features, we extracted the fundamental frequency (F0), the minimum and the maximal value of it (Fmin, Fmax), the first two formants (F1, F2), and the intensity (I). Table 5.12 presents the results of all the interactions and their descriptive statistics. All features were also extracted from the voice of nonhuman agents.

F0 represents the main frequency used for the transmission of speech and can be related to pitch. Although the frequency is directly related to gender, we took the average value to compare the three interactions. There is a significant difference among the three interaction ($F(2, 52) = 33.953$, $p < 0.001$, $\eta^2 = 0.566$), specifically between H-H and H-VH ($p = 0.035$) and marginally between H-H and H-N ($p = 0.045$). As we

expected, we found an interaction between the average value and the variable of the gender ($F(2.52) = 30.685$, $p < 0.001$, $\eta^2 = 0.541$). The average value for the women's voice is 180.651 ± 3.724 Hz whereas for the men's is 128.997 ± 2.150 Hz and thus, the difference between them is 52.655 Hz ($p < 0.001$). There was also a difference of 12.404 Hz between the two ethnicities, but it seems that there is no significance for it. Furthermore, we conducted the comparison between humans' and agents' reactions, and we saw that the average value of Nadine's voice was equal to the women's voice but the one of Nicole was significantly higher. The difference between them and the average human value was significant ($F(2,54) = 115.623$, $p < 0.001$, $\eta^2 = 0.811$) with a significance between every pair of interactions.

Given that the differences in frequencies are expected, as the anatomy between women and men is different, we were interested mainly in the changes during the procedure and among the interactions. That is, to assess how and if we adapt to each thematic area and to see if we are affected by our interlocutor. To complete this, we examined the *minimum* and the *maximum* of the F0 range to see the extent of our pitch. It is of interest that the minimum frequency is almost the same for the three interactions and both genders. There is no significant difference among the interactions but there is one between humans and agents ($F(2,54) = 67.593$, $p < 0.001$, $\eta^2 = 0.715$). Contrariwise, the maximum value of F0 (Fmax) found to be significant among the interactions ($F(2,54) = 12.845$, $p < 0.001$, $\eta^2 = 0.322$) and post hoc comparisons showed differences between all pairs with $p < .001$. The difference between the genders was also found to be significant ($p < .001$). Finally, we found a significant difference between the agents and the human ($F(2,54) = 27.747$, $p < 0.001$, $\eta^2 = 0.593$).

Likewise, we examined the first two formants F1 and F2. Formants are frequency peaks in the spectrum of the acoustic resonance of the human vocal tract [251]. Regarding the F1, there is a significant difference among the three interactions ($F(2,54) = 4.283$, $p = 0.019$, $\eta^2 = 0.137$), and as the post hoc tests indicated, this difference is between H-H and H-VH ($p = 0.021$). No significant difference was reported between humans and agents. F2 depends on the shape of the mouth and the oral cavity and thus, unlike F1, there was a significant difference between male and female participants ($p = 0.024$). There is a significant difference among all interactions ($F(1.26, 34.18) = 5.854$, $p = 0.015$, $\eta^2 = 0.178$), specifically between H-H and H-VH ($p = 0.003$). Both agents presented similar values and the comparison with the humans' values showed a significance ($F(2,54) = 12.583$, $p < 0.001$, $\eta^2 = 0.318$) between H and N ($p < 0.001$) and, H and VH ($p = 0.002$).

The last frequency-related feature is the one of the intensity, or in other words, volume. There was a great statistical significance among the three cases ($F(2,54) = 113.454$, $p < 0.001$, $\eta^2 = 0.808$), which was confirmed by the post hoc tests: H-H and H-N ($p < 0.001$), H-H and H-VH ($p < 0.001$) and, H-N and H-VH ($p < 0.001$). For the human-agent comparison, we also got a significance difference ($F(2,54) = 39.607$, p

<0.001, $\eta^2 = 0.595$), specifically for H and N (p <0.001) and for H and VH (p <0.001) which means that both agents tended to speak a bit louder than the humans.

Table 5.12 Summary of the descriptive statistics for the frequency-related features for the three interactions and for the comparison human/agents. Nonsignificant values are displayed in grey

Features	Interactions	Mean	SD	p		Agents	Mean	SD	p
F0 (Hz)	H-H	137.710	18.176	<0.001		NA	180.151	10.553	<0.001
	H-N	144.316	29.600						
	H-VH	144.456	28.817			VH	193.689	10.915	
Fmin (Hz)	H-H	103.719	11.339	=0.331		NA	132.119	16.132	<0.001
	H-N	106.306	21.134						
	H-VH	102.131	14.737			VH	147.156	13.397	
Fmax (Hz)	H-H	187.537	30.494	<0.001		NA	224.033	15.979	<0.001
	H-N	212.710	44.942						
	H-VH	220.310	49.212			VH	239.566	32.425	
F1 (Hz)	H-H	586.832	37.287	=0.019		NA	594.287	45.120	=0.521
	H-N	602.247	27.488						
	H-VH	606.503	24.368			VH	587.325	46.833	
F2 (Hz)	H-H	1785.901	78.427	=0.015		NA	1883.986	81.597	<0.001
	H-N	1742.359	81.141						
	H-VH	1726.560	65.451			VH	1889.870	83.645	
I (dB)	H-H	39.623	1.834	<0.001		NA	45.861	3.023	<0.001
	H-N	43.695	3.368						
	H-VH	47.550	4.079			VH	45.238	3.407	

- Acoustic features

Lastly, as acoustic features, we extracted the jitter, shimmer, Harmonics-to-Noise Ratio (HNR), and pulses. Jitter and shimmer have successfully been used in describing vocal characteristics. Table 5.13 summarizes the results for the three interactions and their descriptive statistics.

Jitter (J) refers to frequency perturbation which can imply irregularities in the duration of the signal [252]. We found no significant difference among interactions. However, we found a significant difference between Nadine, Nicole, and the humans (F (2.54) = 39.069, p <0.001, $\eta^2 = .0591$) and pairwise comparisons specified it between every pair of interactions. The lower the value of the J, the better. We see that the J of

agents is lower compared to humans' response. This is normal, as the human factor and human emotions can affect the human voice compared to the programmed voice of the robot and the avatar.

Shimmer, on the other hand, refers to amplitude perturbations. The difference among the three interactions was significant ($F(2,54) = 30.310$, $p < 0.001$, $\eta^2 = 0.529$) and post hoc analysis indicated the difference between H-H and H-N ($p < 0.001$) as well as H-H and H-VH ($p < 0.001$). Although the acoustic features we have chosen are dependent on the human anatomy, we also measured the shimmer for the two agents so that we can conduct the comparisons. The difference between them was found to be significant ($p < 0.001$) as well as their difference with the humans ($F(2,54) = 69.719$, $p < 0.001$, $\eta^2 = 0.721$). Although jitter and shimmer are normally measured in a steady voice for each vowel separately, we used the average for the whole voice duration to serve better our purposes.

Table 5.13 Summary of the descriptive statistics for the acoustic features for the three interactions and the comparison of human/agents. Nonsignificant values are displayed in grey.

Features	Interactions	Mean	SD	p		Agents	Mean	SD	p
Jitter (%)	H-H	1.001%	0.081%	=0.541					
	H-N	1.002%	0.093%			NA	0.676%	0.023%	<0.001
	H-VH	0.903%	0.101%			VH	0.488%	0.021%	
Shimmer (%)	H-H	13.78%	1.04%	<0.001					
	H-N	11.84%	2.00%			NA	17.772%	1.016%	<0.001
	H-VH	12.28%	1.58%			VH	15.775%	1.390%	
HNR	H-H	8.314	0.812	<0.001					
	H-N	9.583	1.692			NA	6.423	0.770	<0.001
	H-VH	9.475	1.503			VH	8.150	1.097	
Pulses	H-H	110	13	=0.438					
	H-N	114	23			NA	157	17	<0.001
	H-VH	113	22			VH	167	19	

HNR describes the degree of acoustic periodicity, which means that portrays the relationship of two components: the periodic component and the noise [253]. It is usually used to diagnose voice pathological disorders but, in our case, it can be used to demonstrate any kind of perturbation. It is also mentioned that gender and age can affect its value [253]. We found a significant difference among all interactions ($F(2,54) = 29.466$, $p < 0.001$, $\eta^2 = 0.522$) and post hoc comparisons specified the difference between H-H and H-N and, H-H and H-VH ($p < .001$). The comparison between humans and agents gave us also significant results ($F(2,54) = 35.772$, $p < 0.001$ and $\eta^2 = 0.570$) with post hoc analysis to indicate that the differences are between H and both N and VH. We can see that Nicole's value is similar to the human's one.

Lastly, the pulse implies the rhythm of the speech. We found no noteworthy differences among the interactions, not even for the questions. The difference between humans and agents was clearly significant ($F(2,54) = 100.597$, $p < 0.001$, $\eta^2 = 0.788$), and our post hoc analysis showed us the exact differences between H and NA ($p < 0.001$) and between H and VH ($p < 0.001$).

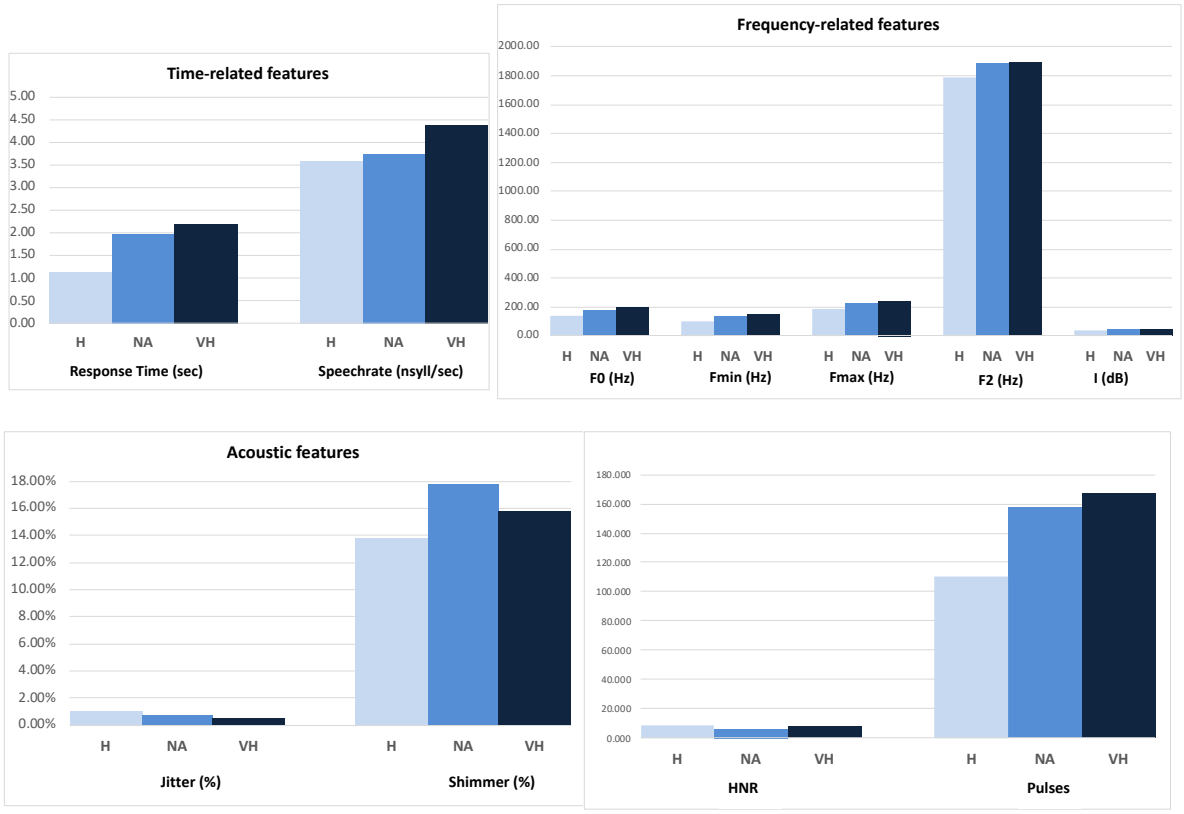


Figure 5.13 Comparison of the features between humans' and agents' responses. Only the features with a significant difference are presented.

Figure 5.13 depicts the features that presented a significance between humans and agents.

To conclude, we notice that participants changed their reactions based on the nature of the questions but apparently, their interlocutor also affected their choices. During H-H interaction, the questions regarding the work environment and the salary elicited the maximum level of participants' responses, meaning that standard, formal questions provoke more stress while interacting with another human. On the contrary, more personal questions, like weakness and hobbies brought out the minimum values of human vocal reactions. HRI followed the same pattern with professional achievements and suitability to draw out more intense human reactions and the imagination of someone's self in five years (yourself_5) the least. The full description of the questions is shown in Table 1. However, H-VH interaction acted differently with the question of weakness to have the highest values of voice features and the one of the work environment to have low values.

Emotion recognition

The results from the emotion recognition are in line with the emotions extracted from the body movements. As we see in Figure 5.14, the dominant emotion during H-NH interaction is sadness. Surprisingly, participants felt calmer and less fear while interacting with the digital human. They presented the same degree of fear while both H-NA and H-H, with no significant difference between them. But they seemed happier while interacting with another human.

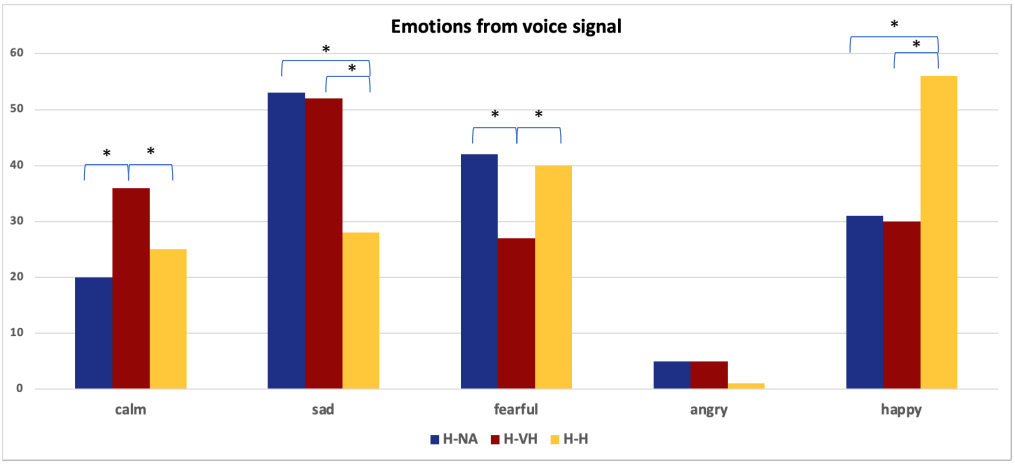


Figure 5.14 Emotions extracted from the voice signal for the three interactions. Significant differences are shown with an asterisk. Y axis represents the amount of time each emotion was recognized as dominant in each interaction.

ML model

Taking full advantage of our voice signal, we also developed a pipeline that successfully separates our data into 3 clusters, proving that from voice data we can actually distinguish whether somebody is talking to a Human, a Robot, or a Virtual Human. The pipeline consists of the LDA dimensionality reduction method and the KNN classifier, as described in the methodology. The former transforms the data to a latent space where they are more easily separable, and the latter finds the best function to fit over the data. The final results of our model are depicted in the matrix of Figure 5.15.

For more details on the performance of the final model, we generated a confusion matrix. More specifically, a confusion matrix gives insights regarding the type of errors being made by the classifier. To generate the confusion matrix we used 10-Fold CV. On each iteration, we normalized all correct and incorrect predictions per class. Subsequently, all values were averaged. Finally, an averaged confusion matrix was produced.

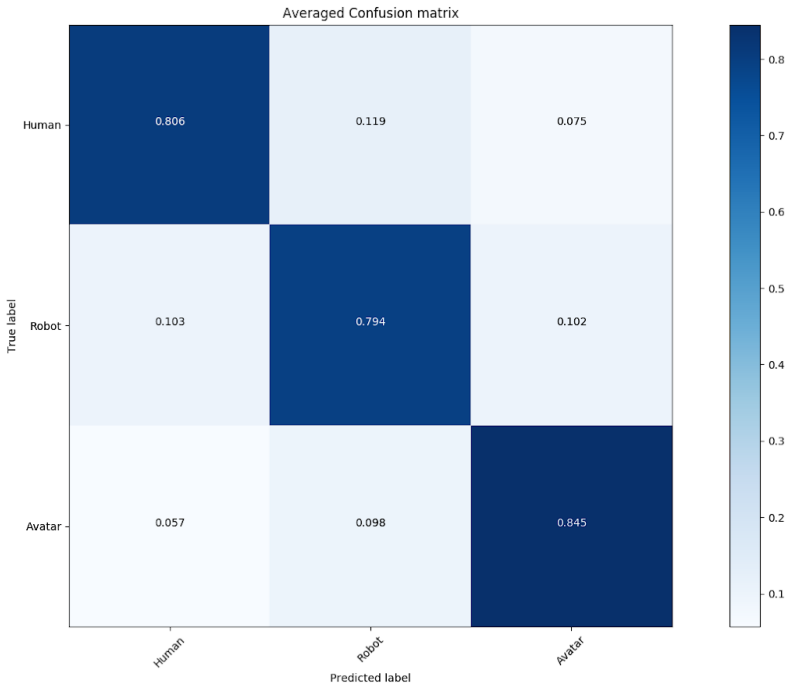


Figure 5.15 Confusion matrix

Lastly, Figure 5.16 shows an approximation of the decision regions of the KNN classifier over the whole dataset. The decision region helps the understanding of how the classifier has decided to divide the input feature space by class label.

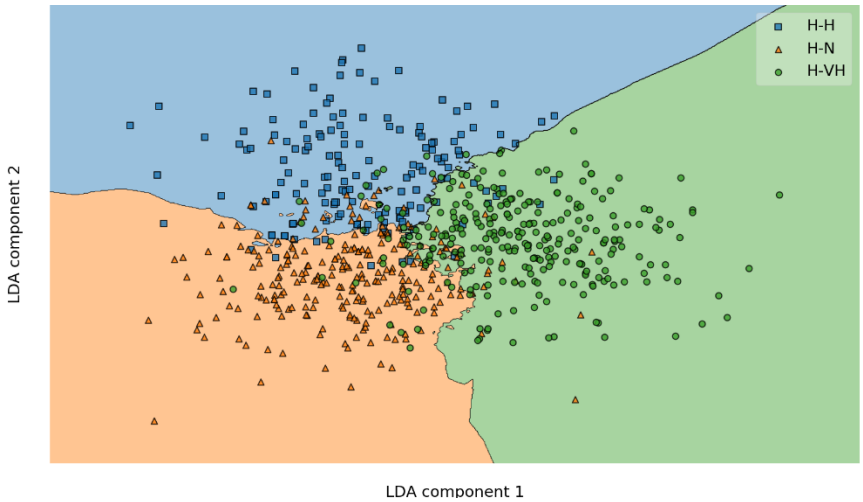


Figure 5.16 Results from the KNN classification (decision boundaries) of the two LDA voice features in the three classes (Human, Robot, Virtual Human (Avatar)) with K = 15.

To find the optimal K for the KNN classifier, a 10-Fold CV was employed. The value of K=15 delivered the most satisfying results. The results of the final model are shown below:

- Precision: 0.818 ± 0.044
- Recall: 0.822 ± 0.043
- F1-score: 0.816 ± 0.044

Psychometric data

Lastly, our results were completed by the subjective psychological reports of our participants via the questionnaire. A clear pattern of differences between positive and negative emotions emerges in all three conditions. Positive emotions are the strongest, negative emotions are the least strong, and surprise is in between, for each condition. Table 5.14 summarizes the results for all the emotions and the post hoc tests using the Bonferroni correction.

Table 5.14 Differences in the strength of Emotions per conditions H-NA, H-VH and H-H

	H-NA		H-VH		H-H	
	M (SD)		M (SD)		M (SD)	
Interested	3.80	$(0.88)^{a}$	3.00	$(1.06)^{b}$	3.82	$(0.84)^{a}$
Confident	3.75	$(0.78)^{a}$	3.70	$(0.97)^{a}$	4.08	$(0.76)^{a}$
Active	3.38	$(0.87)^{a}$	2.92	$(0.86)^{b}$	3.90	$(0.93)^{a}$
Inspired	3.35	$(1.03)^{a}$	2.60	$(0.98)^{b,c}$	3.18	$(0.87)^{b}$
Concentrated	3.20	$(1.02)^{a}$	3.00	$(1.01)^{b,c}$	3.62	$(0.95)^{a,b}$
Surprised	2.98	$(1.14)^{b}$	2.18	$(1.17)^{c,d}$	1.92	$(0.94)^{c}$
Nervous	2.25	$(1.01)^{c}$	1.88	$(0.94)^{d,e}$	1.92	$(0.94)^{c}$
Tired	2.10	$(1.17)^{c, d}$	1.82	$(0.96)^{e}$	1.60	$(0.87)^{c,d}$
Shy	1.82	$(0.78)^{c, d}$	1.38	$(0.67)^{e, f}$	2.08	$(1.12)^{c,d}$
Upset	1.72	$(1.15)^{c, d}$	1.38	$(0.70)^{e, f}$	1.22	$(0.62)^{d,e}$
Afraid	1.55	$(0.71)^{d}$	1.42	$(0.64)^{e, f}$	1.38	$(0.67)^{c,d,e}$
Ashamed	1.42	$(0.59)^{d}$	1.32	$(0.62)^{e, f}$	1.45	$(0.75)^{c,d,e}$
Sad	1.42	$(0.75)^{d}$	1.18	$(0.45)^{f}$	1.12	$(0.40)^{d,e}$
Rejected	1.38	$(0.70)^{d}$	1.60	$(1.01)^{d,f}$	1.25	$(0.54)^{d,e}$

Note: Means in the same row with different letters are significantly different from one another based on post hoc t-tests with the Bonferroni correction ($p<.05$). The absence of letters signifies no statistical significance.

Examining each emotion separately in each condition, we found that there were no differences between the three conditions in Afraid, $F(2, 78) = 1.266$, $p = 0.889$, $\eta^2 = 0.03$, Nervous, $F(2, 78) = 2.939$, $p = 0.059$, $\eta^2 = 0.07$, Ashamed, $F(1.475, 57.540) = 0.745$, $p = 0.441$, $\eta^2 = 0.02$, Rejected, $F(1.742, 67.922) = 2.991$, $p = 0.064$, $\eta^2 = 0.07$, and in Calm, $F(1.699, 66.263) = 0.987$, $p = 0.367$, $\eta^2 = 0.02$. Note that for the comparisons

in Nervous, Ashamed, Rejected and Calm, a Greenhouse-Geisser correction was applied due to violations of the sphericity assumption.

However, Repeated measures ANOVAs exhibited that there were differences between the three conditions in Sad, $F(1.433, 55.884) = 5.521$, $p = 0.013$, $\eta^2 = 0.12$, Upset, $F(1.370, 53.418) = 8.087$, $p = 0.003$, $\eta^2 = 0.17$, Tired, $F(1.75, 68.07) = 5.106$, $p = 0.011$, $\eta^2 = 0.12$, Shy, $F(2, 78) = 10.926$, $p < 00.001$, $\eta^2 = 0.22$, in Surprised, $F(2, 78) = 14.381$, $p < 0.001$, $\eta^2 = 0.27$, in Interested, $F(2, 78) = 15.959$, $p<0.001$, $\eta^2 =0.29$, Active, $F(2, 78) = 13.857$, $p < 0.001$, $\eta^2 =0.26$, in Inspired, $F(2, 78) = 10.074$, $p < 0.001$, $\eta^2 =0.20$, Confident, $F(2, 78) = 5.540$, $p = 0.007$, $\eta^2 =0.12$, Concentrated, $F(2, 78) = 6.130$, $p = 0.003$, $\eta^2 =0.14$. Note that for the comparisons in Sad, Upset and Tired, a Greenhouse-Geisser correction was applied due to violations of the sphericity assumption.

Figure 5.17 shows the dominant emotions during H-NA and H-H, respectively, that presented statistically significant differences with the other two conditions. We see that participants are significantly more surprised during H-NA interaction as well as sad and upset. We have categorized the emotion of surprise under the negative emotions but, given its subjective nature, we could also accept that it could work as a positive one. Remarkably, users were found to be more inspired during H-NA interaction and their interest was equal with the one in H-H but apparently, the process of interacting with nonhuman agents was more tiring. We could also mention that the level of nervousness was noted higher during H-NA interaction, without significant difference from the other two conditions though. On the other hand, while interacting with another human, participants were found to feel more confident, concentrated, and active but in the meantime shyer.

Figure 5.17 The dominant emotions in H-NA and H-H respectively that presented statistically significant differences with the other conditions.

Lastly, we examined the differences in the perception of the participants towards the agents. Repeated measures ANOVAs exhibited that there were differences between the three conditions in Sociable, $F(2, 78)$

= 32.564, $p < 0.001$, $\eta^2 = 0.46$, Friendly, $F(2, 78) = 22.05$, $p < 0.001$, $\eta^2 = 0.36$, and Hostile, $F(1.687, 65.796) = 14.761$, $p < 0.001$, $\eta^2 = 0.28$. We should note that for the comparisons in Hostile, a Greenhouse-Geisser correction was applied due to violations of the sphericity assumption.

Table 5.15 and Figure 5.18 summarize the comparison of the scores and the post hoc tests using the Bonferroni correction.

Table 5.15 Differences in the human perception towards the agents between the conditions H-NA, H-VH, and H-H

	NA		VH		H	
	M (SD)		M (SD)		M (SD)	
Sociable	3.38	$(0.92)^a$	2.62	$(1.00)^b$	4.22	$(0.77)^c$
Friendly	3.78	$(0.80)^a$	3.08	$(0.92)^b$	4.18	$(0.93)^c$
Hostile	1.72	$(0.75)^a$	1.90	$(1.03)^a$	1.20	$(0.40)^b$

Note: Means in the same row with different letters are significantly different from one another based on post hoc t-tests with the Bonferroni correction ($p<.05$)

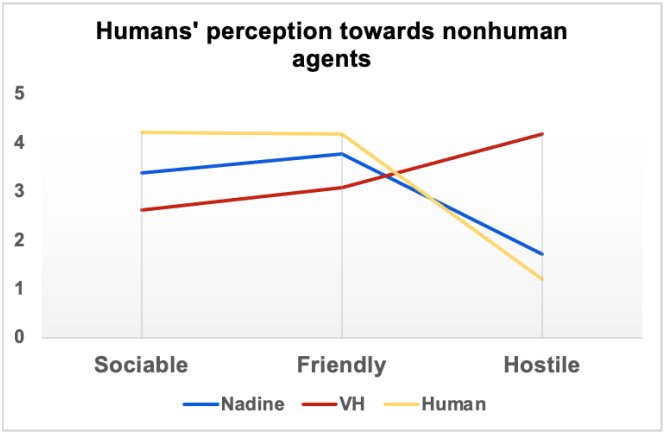


Figure 5.18 The scores of participants' perception towards all the agents.

Reactions' correlations

We conducted a Pearson correlation among all features and all modalities for each interaction. Firstly, we examined each modality with itself. Thus, for H-H, regarding the body movement, we found high positive correlations between all body joints with the upper body (spine base, neck, head) (Pearson's coefficient $r > .700$). However, no correlation was found between the muscles measured through the EMG, except from the two sides of the thenar ($r=0.533$). Regarding voice features, a high correlation between the F0 and the PL ($r=0.781$), among all Fs ($r >0.688$) and between Fmin and Response time ($r=0.688$). Combining all the data, we found an interesting negative correlation between the upper body joints and the gamma band of the PF brain area. The same band was found also to have a positive correlation with the frequency of the

voice. Then, the alpha band of the CP and P areas was found to have a low correlation with the left side of the upper body. Correlations between body joints and brain areas are summarized in Table 5.16. Moreover, as expected, the body joints of the torso (spine, neck, and head) were found to have a high correlation with all muscles. Lastly, an interesting negative correlation was found between the intensity of the voice and the body joints of the arm (r >0.700) as well as with the biceps muscle (r =-0.721), indicating that the volume of the voice is related to arm movements.

During H-NA, we found higher positive correlations between the body joints compared to the H-H (r equal to up to .954) but between muscles, we had only a positive correlation between the two sides of the thenar (r=0.562). Higher correlations were also found between the body joints of the upper body and all muscles, as shown in Table 5.17. PL is positively correlated with F0 (r =0.858) and VB with total duration (r=0.681). Voice frequency was found to be correlated with the alpha band of the CP area as well as the beta band of the P and PF areas. In this interaction, we found a medium correlation between almost all body joints (mainly the upper body) with the frontal alpha band and the upper body joints with the temporal alpha band.

Table 5.16 Pearson Correlation Coefficient r between body joints, voice frequency, and brain states, for H-H and H-NA interactions. * Correlation is significant at 0.05 level, ** Correlation is significant at 0.01 level (2-tailed)

	H-H				**H-NA**					
	Pre-g	**CP-a**	**P-a**	**T-a**	**Pre-b**	**Fro-a**	**CP-a**	**P-b**	**T-a**	**Occ-a**
F0	0.653**	-0.087	0.059	-0.101	-0.515**	0.089	0.590**	0.697***	0.066	-0.184
Neck	-0.544*	0.296	0.321	0.358	0.321	0.626**	0.411	0.382	0.613**	0.557*
Head	-0.546*	0.296	0.204	0.211	0.187	0.552*	0.320	0.260	0.507*	0.502*
Spine Base	-0.547*	0.337	0.379*	0.372	0.407	0.635**	0.407	0.371	0.578*	0.520*
Shoulder L	-0.620**	0.282	0.270	0.285	0.199	0.628**	0.374	0.302	0.552*	0.514*
Elbow L	-0.544*	0.512*	0.502*	0.507*	0.341	0.665**	0.276	0.291	0.638**	0.577*
Wrist L	-0.391*	0.521*	0.506*	0.560*	0.311	0.658**	0.199	0.253	0.668**	0.635**
Hand L	-0.369	0.534*	0.516*	0.563*	0.289	0.668**	0.243	0.287	0.690**	0.670**
Shoulder R	-0.663**	0.141	0.179	0.161	0.103	0.613**	0.319	0.364	0.533*	0.514*
Elbow R	-0.580**	0.186	0.161	0.413	0.170	0.631**	0.366	0.255	0.579**	0.477*
Wrist R	0.532**	0.460	0.281	0.219	0.232	0.568*	0.210	0.175	0.558*	0.321

Note: Pre-g: Prefrontal gamma, Fro-a: Frontal a, CP-a: CentroParietal a: P-a: Parietal a, P-b: Parietal beta, Occ-a: Occipital alpha

Lastly, for the H-VH, we found a correlation between the two sides of the abdominal muscles (r=0.533) and of the thenar muscles (r=0.683). Total duration is again correlated with VB (r = 0.560) but lower than in H-NA and PL with the F0 (r=0.750). Correlations between the body joints of the Kinect data were found lower compared to H-NA, as well as between the upper and the lower body. The correlations between the muscles and the body joints of the upper body were also lower. Some muscles (Abs and Thenar) were found also to be correlated with the lower body (Knee, Ankle and Foot) with r up to 0.761.

Table 5.17 Pearson Correlation Coefficient r between upper body joints and muscles, for all interactions * Correlation is significant at 0.05 level, ** Correlation is significant at 0.01 level (2-tailed)

	H-H							
	Biceps R	**Biceps L**	**Trapez R**	**Trapez L**	**Thenar R**	**Thenar L**	**Abs R**	**Abs L**
Spine Base	0.764**	0.789**	0.756**	0.772**	0.531**	0.598**	0.778**	0.771**
Neck	0.864**	0.907**	0.888**	0.885**	0.654**	0.664**	0.893**	0.906**
Head	0.886**	0.889**	0.965**	0.961**	0.689**	0.650**	0.831**	0.794**
	H-NA							
Spine Base	0.913**	0.835**	0.892**	0.863**	0.602**	0.654**	0.879**	0.834**
Neck	0.921**	0.945**	0.925**	0.947**	0.713**	0.744**	0.948**	0.946**
Head	0.885**	0.927**	0.965**	0.967**	0.767**	0.752**	0.821**	0.807**
	H-VH							
Spine Base	0.602**	0.609**	0.500*	0.609**	0.469**	0.519**	0.702**	0.635**
Neck	0.827**	0.804**	0.798**	0.804**	0.278	0.088	0.869**	0.765**
Head	0.745**	0.839**	0.921**	0.839**	0.223	0.213	0.665**	0.579*

Role of gender and ethnicity

Although our data were not balanced, we controlled possible interactions with gender or ethnicity and finally, we evaluated only differences with high statistical significance. The value of partial eta also helped us to assess the role of the sample size. Thus, throughout our modalities, we found some very interesting differences in some features regarding both gender and ethnicity.

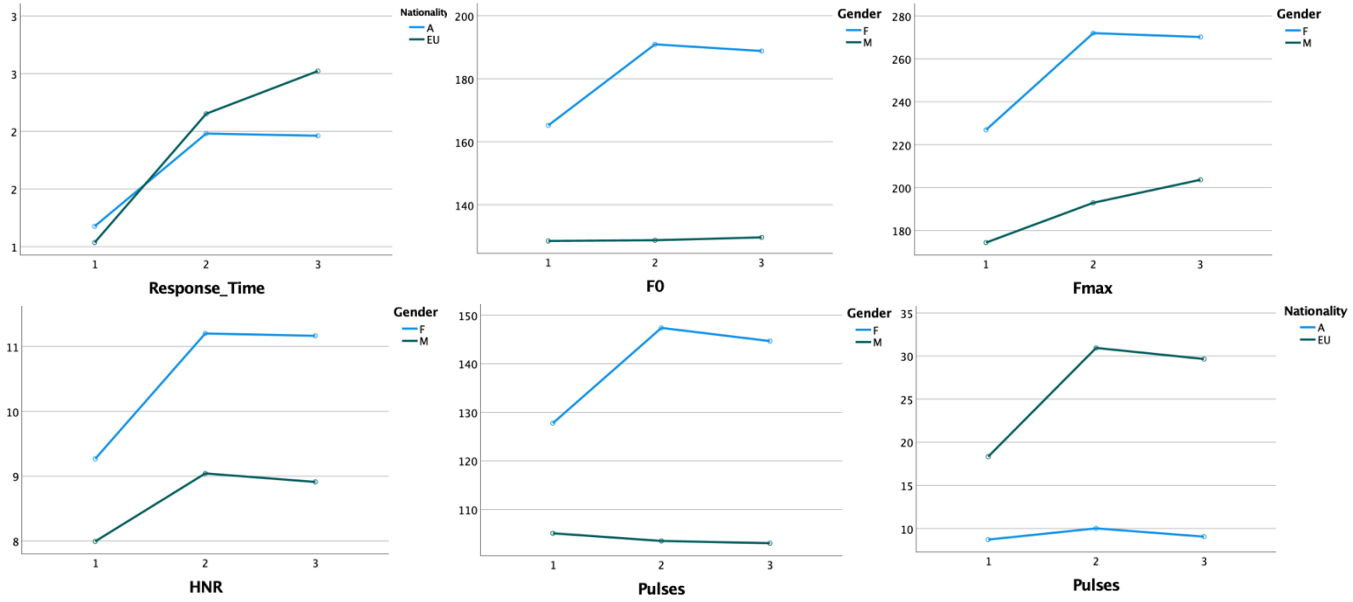


Figure 5.19 The significant differences found in features in function of the gender and/or ethnicity. 1 = H – H Interaction, 2 = H – N Interaction, 3 = H – VH Interaction

Figure 5.19 illustrates such differences for the audio modality. First of all, Asian participants tended to respond faster while interacting with nonhuman agents compared to Europeans. However, the latter tend to speak more. This is something we also observed during the experiments. In general, there were features where we found no significant difference during our statistical analysis but during the experiments, our observations gave us a different feeling. For example, Asian participants seemed to make longer pauses while speaking compared to the European ones whereas the latter seemed to speak louder. Equally, the difference in intensity between the two ethnicities was not significant but during the experiments, we had the impression that Europeans were speaking louder which may be a result of their higher expressivity.

Regarding frequency, we got, as expected, a difference between genders. However, the Fmin presented no difference. Shimmer and HNR were also influenced by gender, with female participants showing a lower value of S and a higher of HNR, which is also in line with the study of Yumoto who found an association between HNR and hoarseness; the lower the value of HNR, the higher the level of hoarseness [254]. Women are usually expected to present less hoarseness in their voice, compared to men. Lastly, although for pulses we found no significant differences among the interactions, gender and ethnicity gave us different results. Europeans had more pulses than Asians in their voice and women more than men.

Regarding our Kinect data, no significant difference has been noticed. However, of interest is that in our gender analysis, although we found nothing significant between the three interactions, we noticed that the value of each body part between females and males is the same during H-NA interaction whereas there is always a difference during the other two interactions.

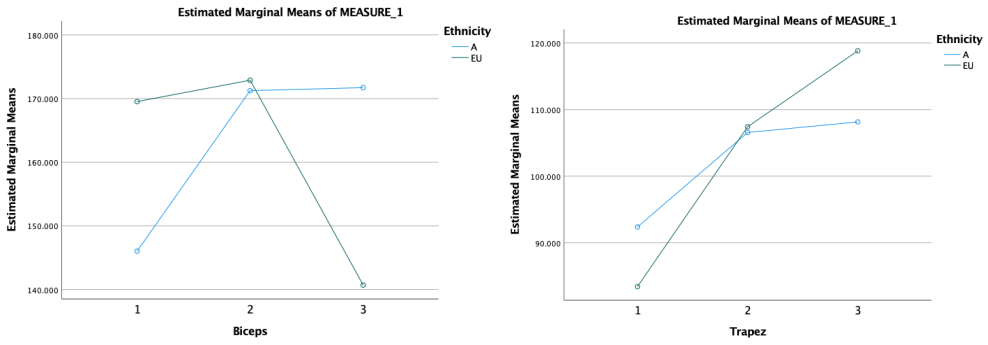


Figure 5.20 Mixed ANOVA with ethnicity as the between-subject variable for Biceps and Trapezius for the three interactions. 1. H-H interaction, 2. H-NA interaction and 3. H-VH interaction.

Gender and ethnicity analysis for EMG data gave us no further significant results except the muscle of the arm, the Biceps. Ethnicity was found to have an interaction with this muscle ($F(2,74) = 4.108$, $p=0.020$, $\eta^2= 0.100$) between H-H and H-NA ($p=0.003$) as well as H-NA and H-VH ($p<0.001$). Moreover, we noticed that in general, for all muscles except the abdominals, the human responses during H-NA had always

approximately the same value between males and females, as well as Asians and Europeans. Figure 5.20 confirms this via an example of Biceps and Trapezius from the ethnicity analysis.

5.2.5 Discussion

In this study, we performed an in-depth analysis and comparison between the natural human-human interaction and the human-computer and human-robot ones. Our goal was to find and bridge the gap between human and nonhuman interactions, extracting humans' features that can help us understand human behavioral and emotional processes. The features can act as a tool for a more successful and fluid collaboration between humans and robots or other kinds of similar technology (i.e digital humans). What is mainly missing from the up-to-date state-of-the-art is the direct comparison of any kind of nonhuman interaction with the original human-human communication and the clarification of human behavioral patterns in the context of both natural and technological frameworks. Towards this direction, our results allowed us to answer adequately our third research question: *How do humans' voice and body react when interacting socially with humans compared to nonhuman agents?*

Human reactions

Brain activity

Regarding brain activity, we estimated the relative energy of each brain state in five brain areas. In the Prefrontal area, the theta state was found to be higher for all interactions, with no difference between them.

In the Frontal area, the energy of the theta state was found to be significantly higher during H-NH interaction, indicating the cognitive load [37]. Increase in theta activity is also associated with the initial learning improvement [39]. Our result is in line with previous works [201] that pointed frontal theta oscillations when interacting with a humanoid, which is also extended to the digital humans in our case. Moreover, increase in the fronto-central theta waves has also been related to detecting prosodic emotional changes [189] and thus, theta band is associated with the perception of emotions through vocal expressions [190]. In this area, we also noticed high energy of the delta frequency band during H-H and H-NA. In general, delta oscillations play an important role in cognitive processes like attention, memory, and decision making and they are focused on frontal, central, and parietal areas, as well as occipital if they are related to emotional processes [191]. Specifically, they have been associated with arousal in posterior brain areas and with valence in anterior brain areas, as well as with surprise [39]. Moreover, delta activity in frontal areas has also been linked to the perception of face recognition related to emotional expressions [192]. The fact that we noticed this energy during H-H and H-NA makes us think that the physical presence of these agents

(human and robot) facilitates the perception of face recognition. Lastly, gamma band was higher only during H-NA. Unpleasant stimuli can trigger effects in gamma range and higher frequencies have also been found to be a reliable indicator of arousal [39].

In the Parietal area, as expected, we found high energy of alpha waves during H-NH interaction which is in line with our previous studies working in HRI or VR as described before. Since our experiment doesn't involve VR but interactions with nonhuman agents, we may assume that it can also be used as an indicator of an H-NH interaction. However, further investigation is needed for the clarification of alpha oscillations' role in this field. Alpha oscillations in the Temporal area are also in line with our previous studies, probably indicating the effort of participants to decipher the voice of the nonhuman agents.

Lastly, in the Occipital area, as we expected, we found high energy of alpha waves during H-NH, which indicates the engagement level of visual attention mechanisms. During H-NH interaction, the brain is in process of a smaller repertoire of non-familiar faces and objects, compared to the H-H one, where we noticed increased beta waves. Higher beta activation can be elicited by dynamic emotional face expressions [190]. Moreover, during H-NH interactions, a medium level of delta band energy was noticed, which can be associated with emotional processes and specifically arousal.

Motion data

Regarding the motion data, combining the results derived from the Kinect and the EMG, we interestingly noticed that in general, the left side of the body had a bigger range of movement as well as the lower part, but the right side had higher intensity and activity of muscles for all interactions. Our results led us to assume that most of our participants were right-handed. We conclude that the left side is more relaxed and thus, the range of the motion can be bigger. Specifically, when comparing H-H with H-NH interactions, the differences in motion between the two sides are smaller but the range of movement is bigger. The latter can be considered as spreading of movement and based on the Laban Movement Analysis (LMA), it reveals a happy emotion [35]. However, we should note that during H-NA the range of movement was bigger compared to H-VH, revealing that participants were more motivated to move in front of a robot than of a digital human. During H-NH interactions, muscles presented higher activity with significantly higher mean frequency compared to H-H, and higher intensity with higher RMS but no significant difference was found between H-NA and H-VH. Specifically for the biceps muscles, we found high Fmean, which means that the muscle was often used, executing movements like supination of the arm or elbow flexion, but with low intensity. In other words, small, sharp movements were executed. This kind of movements can be associated with the emotion of anger [35]. The trapezius muscle follows, which means shoulder and upper back were used, but a medium level of intensity. Given the nature of the movements, i.e elevation of the shoulder, this outcome could signify a lack of comfort. Regarding the thenar muscles, we found lower Fmean but high

levels of RMS, which can be translated as longer movements with higher intensity, like the clenching of the fingers. We will consider this movement compressed and confined and thus, we will relate this movement to the fear [35]. Lastly, results from the abdominal muscles were inconclusive as only the RMS value was found to be slightly significantly higher during H-NA, meaning that participants either were breathing a bit faster or/and they did more intense movements with their trunk.

Vocal behavior

Participants' vocal behavior gave us also interesting results. First of all, we noticed that participants spent more time on their answers when interacting with another human. Moreover, we noticed that the pause duration is lower during H-H, but the number of voice breaks is higher. The response time during H-H was significantly lower, which indicates a better flow in the discussion. For this feature, we noticed that the human response was correlated to the agent's response and we started wondering if our responses are influenced by our interlocutors. To confirm this assumption, we conducted a Pearson correlation for this feature, as shown in Figure 5.21, and we found that the higher the response time of the interviewer the higher the value of the participant as well.

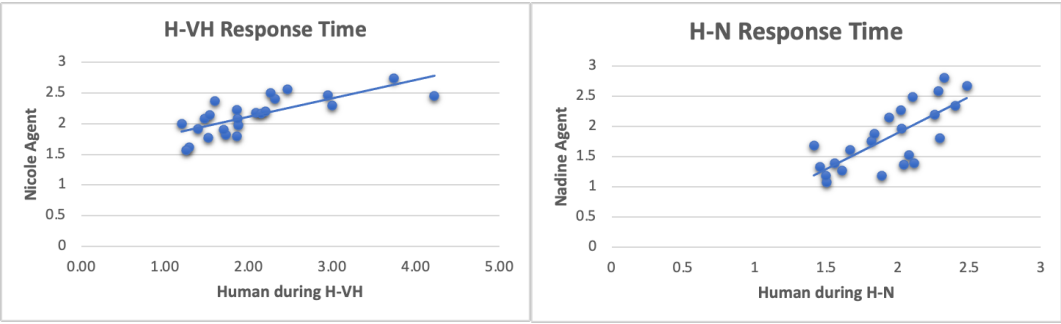


Figure 5.21 Response time between the participant and the agent. Pearson r correlation coefficient is equal to 0.744 and 0.740 from left to right.

The same result we noticed for the speech rate but not with such a strong correlation. The highest value of Sr was detected during H-H which means that we tend to speak faster with other people.

Of interest is the change in the frequency while interacting with agents. To facilitate our purpose, we used the average value of frequency, men and women included. We take into account that both nonhuman agents are females and human agent as well. However, the significant outcome is that frequency is lower when we speak with people and higher when we speak with nonhuman agents. So, we cannot claim that mimicry plays a role for this feature. We also noticed that the lower values of F presented no significant difference, but the maximum values do. Moreover, participants tended to speak louder when interacting with Nadine

or Nicole. That could be a sign of stress, but it can also be linked with the fact that participants felt a higher level of shyness interacting with the human. Shyness can decrease the volume. Apparently, the nature of the agent also played a role In the human responses as e.g participants spoke slower and softer to Nadine compared to Nicole.

Lastly, we examined acoustic features to find any perturbations in the voice. Although jitter gave us no significant differences, shimmer was found higher during H-H interaction. HNR is considered a sensitive index of vocal function. In our study, the highest value of HNR was found for the H-N interaction whereas the lowest one for H-H. As we mentioned before, HNR is related to hoarseness [254] so we can conclude that participants tend to present some hoarseness while interacting with the human.

In summary, when interacting with nonhuman agents compared to another human in a context of a job interview, we tend to give shorter answers with longer pauses and to speak slower. The frequency and the intensity of the voice are higher. Perturbations in amplitude are less and the value of HNR is bigger. Increased levels of frequency combined with slower speechrate and high volume are associated with nervousness and agitation [33]. Moreover, high levels of volume with high frequency are also related to fear [34]. In general, differences in vocal behavior were obvious among the three interactions and that was verified by our classification model.

Psychological data

Lastly, we combined the results from the questionnaire with the emotion recognition from the audio and the movement. The questionnaire is the only subjective measure we have in our study, and it was very compelling to have the opportunity to see how participants evaluated themselves compared to how their body and brain actually reacted. Thus, participants claimed to be more interesting and inspired during H-H and H-NA interactions but in the meantime shier. During H-H interaction, happiness, confidence, and concentration were the dominant emotions whereas during H-NH were mostly negative ones, such as agitation and surprise. The digital human received a medium score for almost all the emotions. On the other hand, voice and motion showed us mainly negative emotions during H-NH interactions, with sadness, fear and even neutral to be the dominant ones. Happiness is always present during H-H interaction, and voice showed us that during H-VH interaction participants felt calmer. H-VH interaction seemed to be more familiar to participants, compared to the H-NA one, as they were found less shy, less fearful, and calmer but also with less interest and inspiration.

Summary

Summing up, brain results showed us that participants tended to deal more with their emotions during H-NH interactions, and especially during H-NA. Interaction with the nonhuman agents seemed to make them

more concentrated, with a higher cognitive load. They used to speak louder and slower, with a higher frequency, longer pauses, and shorter answers. The nature of the agent affected their vocal behavior. Movements had a narrow range but with a big difference between the two body sides and they were more intense, especially during H-NA. In general, the left side presented larger movements but the right-side had higher activation. H-NH interactions provoked mainly negative emotions in participants but the interaction with Nadine also triggered interest and inspiration. Figure 5.22 summarizes these results.

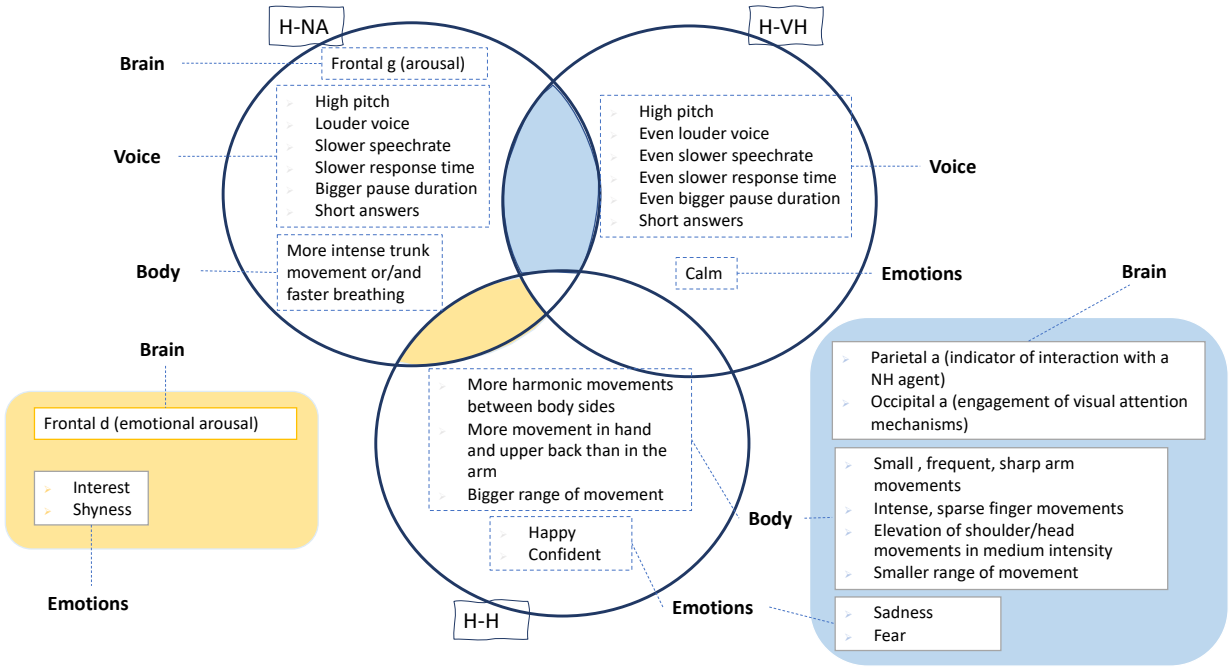


Figure 5.22 Summary of the human reactions per interaction for each modality

To complement the above, we also examined possible correlations among all human reactions. The frequency of the voice and specific body joints was found to be correlated with brain bands and all the examined muscles were found to be correlated with the movements of the head, neck, and low part of the spine. Specifically, during H-H interaction, the voice frequency and the upper right body side correlated positively and negatively, respectively, with the gamma band of the Prefrontal area. As we have mentioned before, the gamma range is triggered by unpleasant or arousing stimuli [38]. The negative correlation could verify the dominance of positive emotions during H-H. However, the upper left body part is positively correlated with the alpha band of the CP, P, and T areas. During H-NA, all upper body joints are positively correlated with the alpha band of F, T and Occ areas. The alpha rhythm becomes coherently engaged in transforming perception to action [192] and thus, this relationship could indicate the concentration and the effort of the participant to proceed with an action. Comparing these two interactions, we verify that during

H-H, the participants felt more familiar and natural and there was a direct activation of the body language. Surprisingly, during H-VH, we found no strong correlations regarding brain activity.

The high positive correlation between the muscles and the head, neck, and spine seems to be affected by the nature of the agent. Correlation is higher during H-NA, which we could explain as increased nervousness. During H-VH, the arm seems to be more independent, which can confirm the emotion of calm extracted from the audio signals.

Role of gender and ethnicity

To complete the answer to our third research question, we need also to reply to the subquestion: *Do gender or ethnicity affect human behavior when interacting with a nonhuman agent?* Although our sample was not balanced, we examined possible differences due to gender or ethnicity and with the help of the partial eta squared which indicated the role of the sample's size, we kept results only with high statistical significance. However, our results didn't give us a lot of insight. Regarding gender, the main differences concern vocal behavior. We noticed that both genders had similar minimum frequency but women presented a higher maximum one, letting us wonder if women have a broader voice frequency range. Moreover, female participants presented lower value of shimmer and higher value of HNR, which is in line with the fact of hoarseness we mentioned above. Lastly, Europeans had more pulses than Asians in their voice and women more than men. Ethnicity control showed us also differences in vocal behavior but also slightly in movement. Asian participants tended to respond faster whereas Europeans to speak more. Moreover, the latter used more their arms when they were interacting with another human but less when interacting with the virtual human. All participants had similar reactions during H-NA interaction.

General Conclusion

Summing up, this study aimed to provide insights regarding human brain activity, motion, vocal behavior and emotional states from a direct comparison between a natural human-human interaction and an interaction with a social robot or an avatar. The scenario was the same for all interactions and it concerns the first phase of a typical job interview. Given the proliferation of the use of non-human agents in professional contexts, such as that of a job interview, we believe that our outcome can help the development and adaptation of technological systems, like job interview systems, as well as future applications of human-robot and human-computer interaction in general. We argue that studying human reactions can provide an understanding and meaningful implications for future research in this direction. This is supported by the fact that 72% of the participants admitted to being less nervous discussing with the social robot or the digital

human, thus demonstrating the advantages of adapting non-human agents to better match the human behavior and the human needs, facilitating a more natural interaction.

5.3 Potential and Acceptance of Social Robots

5.3.1 Experimental design

The *second part* of our study included only the human-robot interaction but under different predefined scenarios:

- **Job Interviewer** (as in part 1)
- **Customer guide** in a shop with electronics, where participants look for a new cellphone. Nadine discussed with them, asking questions regarding possible characteristics participants would like to have and at the end, she proposed a model.
- **Teacher**, where Nadine gave a short lesson regarding climate change. She explained the term of climate change and the current situation, interacting with the participants by asking several questions on the topic.
- **Companion**, where participants were able to interact freely with Nadine on a topic of their choice. She was following the flow of the chosen subject. Non-native English speakers had the opportunity to try speaking in their native language (French, German, Chinese, Hindi)

The purpose of this part is to examine how people react towards robots, how a robot can affect their reactions and what would be the ideal role for them to support. In the first three roles the scenario was predefined whereas in the companion mode participants were interacting freely with the robot. Figure 5.23 shows an example of a participant, interacting with Nadine under the role of customer guide.

Figure 5.23 Example of a participant interacting with Nadine in the role of the customer guide. The participant wants to buy a cellphone and Nadine asks for several characteristics in order to decide which one would be the ideal cellphone for him, according also to his budget.

5.3.2 Data acquisition and analysis

In this part of the experiment, we used only EEG, motion data, and the two questionnaires to reach our goal.

EEG recordings and analysis

EEG was recorded and amplified as in part 1 of the experiment.

The analysis of the EEG data and the processing of the signal were carried in MATLAB. All data were carefully checked for artifacts, like eye blinks or head/body movements. Fast Fourier Transform was applied and then the power spectra were calculated. We examined 5 ROIs including both hemispheres: Prefrontal (Fp), Frontal (F), Parietal (P), Temporal (T) and Occipital (O), as ahown in Figure 5.24, for four brain states: theta (3-7 Hz), alpha (8-12), beta (13-30 Hz) and low gamma (30-42 Hz).

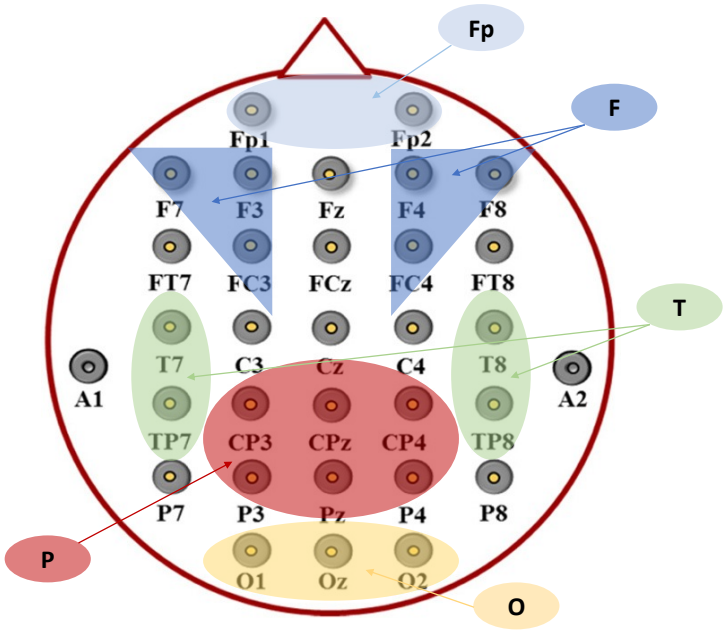


Figure 5.24 The five Regions of Interest (ROIs) used for both parts of the experiment. 23 electrodes were selected according the needs of our research.

Motion data

The analysis of the motion data was done exactly as described in 5.2.3.2, regarding emotion recognition.

Psychometric data

To measure participants' attitudes towards robots in general, before their interactions with Nadine the social robot, we used a slightly adapted version of the NARS questionnaire [255].

Moreover, as in part 1, we used a questionnaire based on the Positive and Negative Affect Schedule [250].

5.3.2 Human perception in HRI under different roles

The second part of our study targets the understanding of human expectations and acceptance of socially intelligent robots. Thus, we urged participants to interact with Nadine under four different scenarios, for four different roles.

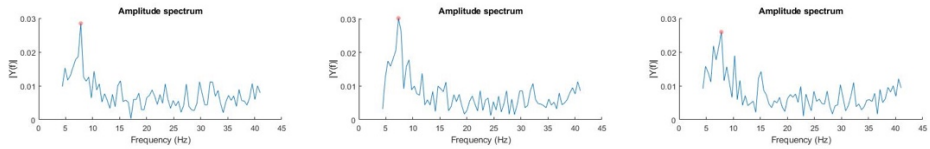
Participant's reactions and preference over four predefined robot roles

Brain activity

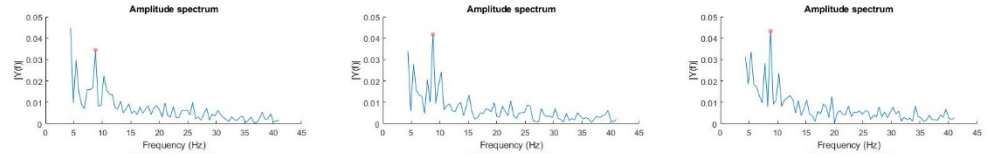
The main purpose of the EEG use in this experiment was to examine how the brain reacts when interacting with a robot under several roles. Consequently, we can assess human perception towards robots, revealing humans' true emotional states and attitudes, and finally decide which role, under the existing technology, is more preferable and more productive.

As we mentioned, we have focused on five brain areas. Thus, starting from the Prefrontal area, we noticed a dominance of theta rhythm for the role of the interviewer whereas for the three other roles an alpha state was maintained. However, there was an increase in the frequency from one role to another (teacher, customer guide and, companion accordingly), which may act as a factor of familiarity. The Prefrontal cortex is completely associated with personality traits, planning of social behaviors, and decision making. Theta state in prefrontal area has recently been associated with spatial working memory [256]. Working memory refers to a temporarily processing of information to cope with complex tasks. So, it is not surprising that a theta state was presented under the interview phase as the users had to reply to answers retrieving several kinds of information. The alpha rhythm shows a strong engagement of the decision making and behavior mechanisms. The theta state corresponding to the interview phase (7.5 ± 1.2 Hz) presents high statistical significance towards the other roles ($p < 0.01$). Among the other roles were alpha state was found, only the teacher (9.7 ± 1.2 Hz) and the companion (11 ± 1.7 Hz) presented a statistical significance ($p<0.05$) which may mean that the increase in the frequency does reveal a factor of familiarity as the users experienced the four roles in a raw.

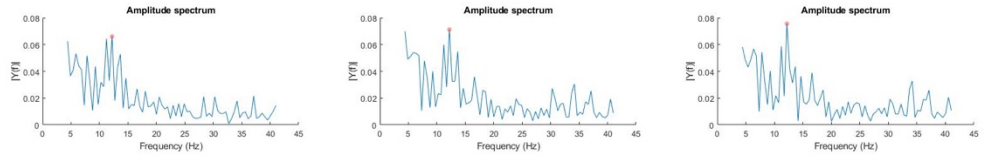
In the Frontal cortex, we noticed the existence of frontal theta oscillations for the role of the interviewer (7.2 ± 1.1 Hz), with a high significant difference from the others ($p<.001$), revealing that the task was demanding [230] and participants had to put a higher cognitive effort to get focused and interacting with Nadine. We remind that Nadine as an interviewer was the first role that participants had the experience with, thus the first contact with the robot. Subsequently, for the second role of the teacher, we noticed the alpha rhythm (12 ± 2.5 Hz) which can be associated with the concentration of the participants. The role of the teacher presented a significant difference ($p<0.05$) with all the other roles as well. For the rest two roles, a low beta state was apparent (14 ± 3.3 Hz for customer guide and 14.3 ± 1.7 Hz for companion) which means that participants had no stress during these interactions and they were mentally alert [251, 257].



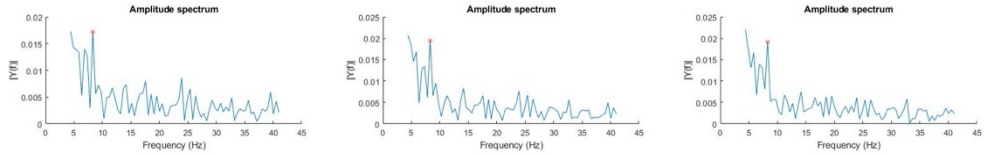
(a) Interviewer



(b) Teacher



(c) Customer guide



(d) Companion

Figure 5.25 The three channels recorded in occipital area showing the dominant brain state for each role in the power spectrum. Nadine's roles are shown in the order they were presented in participants. In the case (a) we found a very low alpha state (7.7 ±

1.1 Hz), whereas for the rest three roles we noticed a clear alpha state with an increase of the frequency for each role. Case (d) concluded with an average value of 11 ± 2.3 Hz.

Regarding the Parietal cortex, alpha rhythm was dominant for the roles of the interviewer (11.8 ± 1.6 Hz), teacher (12.5 ± 1.8 Hz), and companion (10.2 ± 1 Hz), with the latter to present a significant difference from the other two (p =0.003). This result is in line with all our previous outcomes. Alpha rhythm in the parietal cortex is also associated with emotional engagement and this explains the presence of this rhythm for the role of the companion. However, for the roles of the customer guide (13.4 ± 2 Hz), we found a beta state.

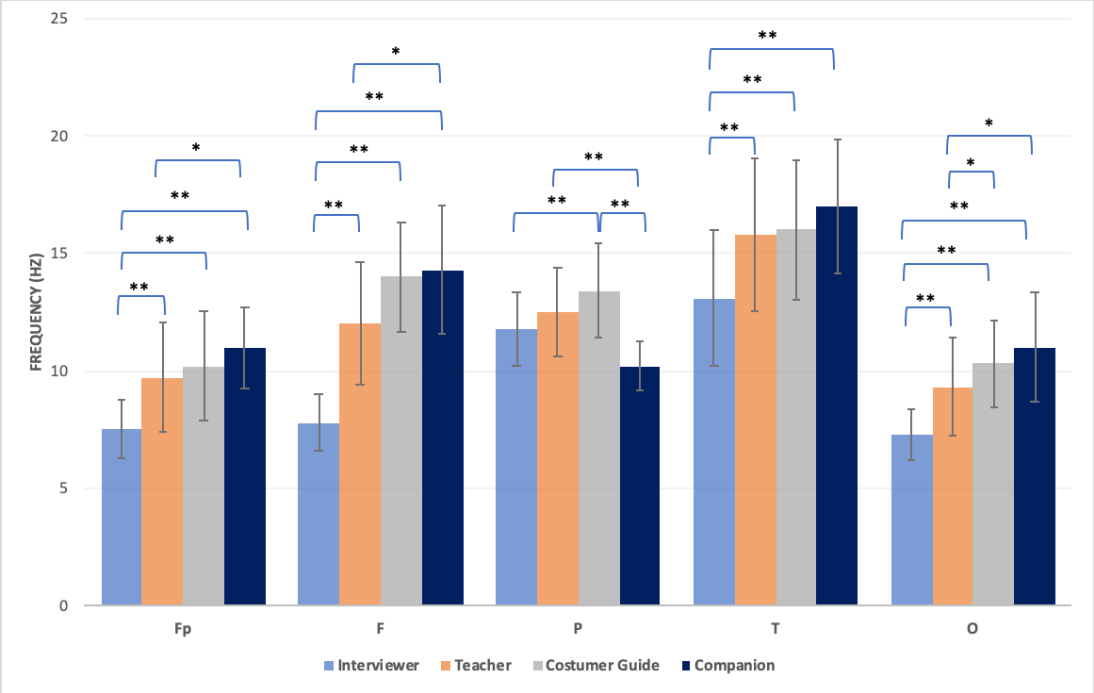


Figure 5.26 The average frequency in the five brain areas for the four roles of Nadine. Statistical significance is shown where existed (* for p<0.05 and ** for p<0.001 that we consider as high significance).

In the Temporal lobe, we found a general beta rhythm with an increase in the frequency in each role. This could reveal that participants had no real difficulty following Nadine's speech. However, the role of the interviewer presented a lower value (13.1 ± 2.9 Hz) and has a highly significant difference (p<0.001) with all the other roles. This means that participants faced the biggest difficulty in deciphering the robot's voice while being interviewed, which may be explained by the fact that this role was the first to interact with and participants had no prior robot experience. The more they interacted with the robot, the more familiar they became with it, in terms of audio processing. For the companion role, we noticed a value of 17 ± 2.9 Hz.

Lastly, in our previous study, we found that the human brain can understand the difference between a human and a robot, even if the robot shares the same physical appearance as the human. In that case, while looking in a humanoid robot the occipital area of the brain is synchronized in an alpha state whereas while we interact with a human, beta state is dominant. In our experiment, we verified this finding by having the alpha rhythm in all the roles. The continuous increase in the alpha state reveals that the more we interact with a robot, the more familiar we become with its appearance. All the roles presented a significant difference between them, except the last two, customer guide and companion. Figure 5.25 presents an example of the EEG activity in this area.

Figure 5.26 presents the average value of frequency for each role in each brain area. Statistical significance is also shown.

Kinect data

We extracted seven discreet emotions through the Kinect recordings of body movements. We examined the 25 body points as described above, divided in 5 body parts and we classified these movements in positive and negative emotions. Figure 5.27 depicts this classification.

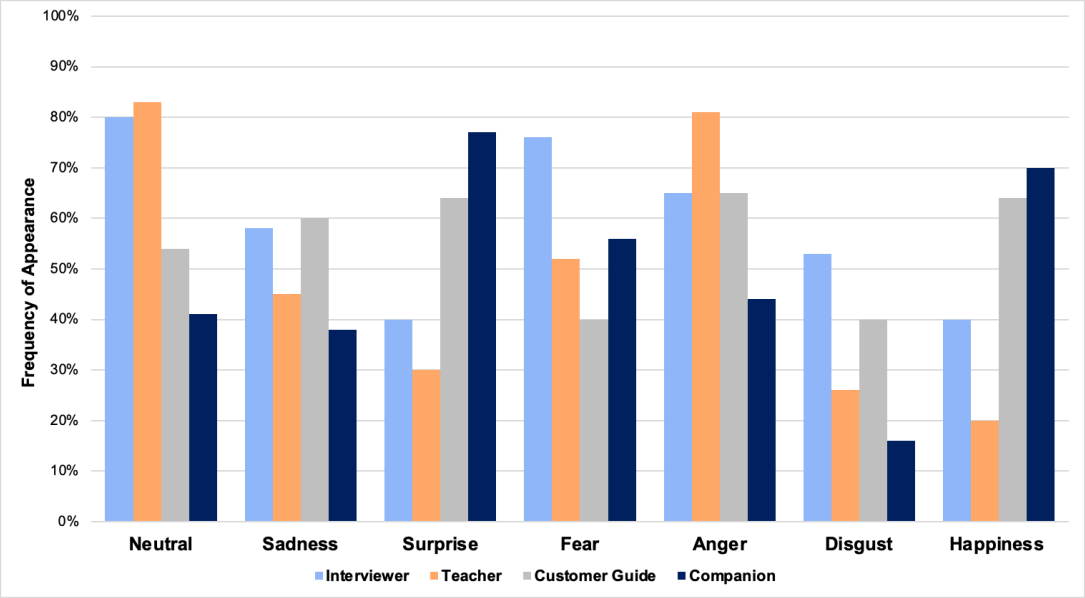


Figure 5.27 Discreet emotions extracted from body skeleton movements through Kinect V2.

Sadness was the most prominent emotion during the role of customer guide, followed closely by the role of interviewer. Apparently, the customer guide elicited high scores in almost all emotions. Companion had the highest rate of happiness, followed by the customer guide and the teacher. The surprise was also the most prominent during the companion and the customer guide roles and the least during the teacher role. Fear was evoked mostly during the interviewer. Anger was presented highly during the teacher role, letting

us wonder if the scenario of the interaction (global warming) played a role. Finally, participants were most disgusted by the interviewer role and the least by the companion role.

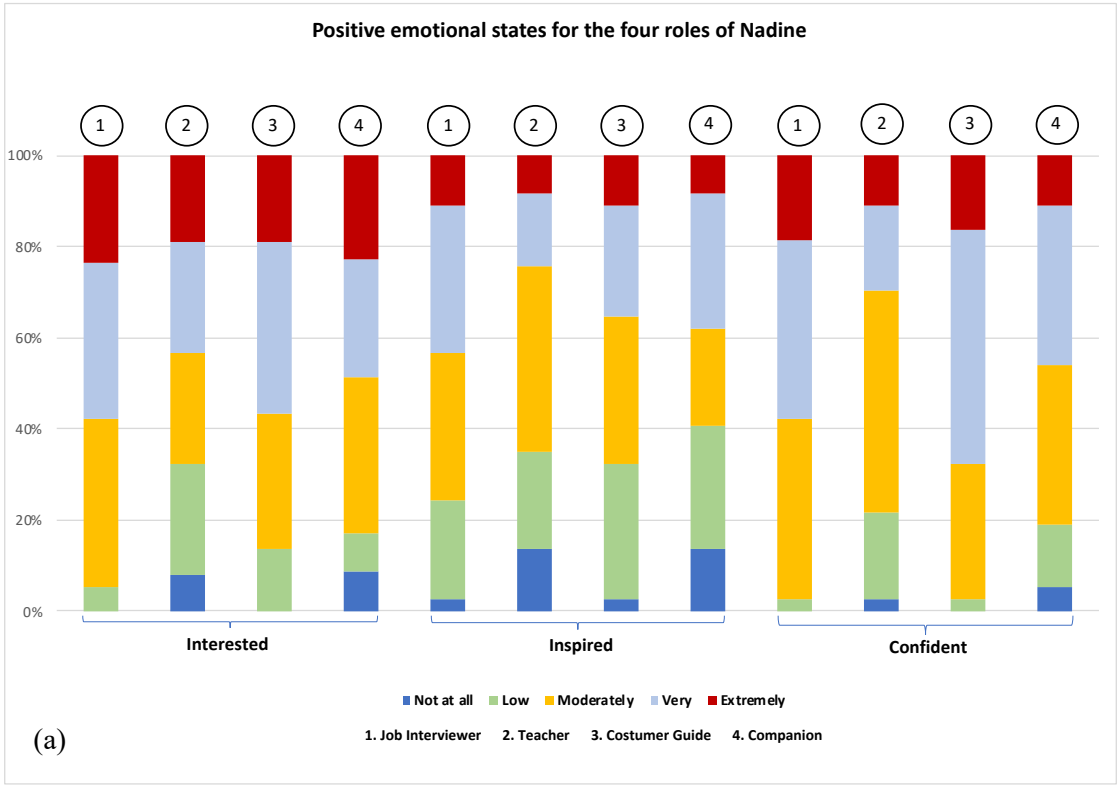
Psychological data

The questionnaire complemented the above providing information regarding the emotions of participants during their interactions and their preference towards the 4 roles.

We noticed no high levels of negative emotions throughout the process. We examined three negative (nervous, shy, and tired) and three positive (interested, inspired, confident) emotions/states. For the positive emotions, a significant difference was found among the roles for each emotion (($F(2,432) = 9.789$, $p < 0.01$) for interest, ($F(2,432) = 7.463$, $p < 0.01$) for inspiration and $F(2,432) = 8.332$, $p < 0.008$ for confident). Specifically, participants were found to be significantly more interested and confident than inspired ($p < 0.001$). This pattern was the same for all roles. Furthermore, the interviewer and the customer guide roles were not statistically significantly different ($p = 0.513$), but the two had higher scores on positive emotions compared to both teacher and companion roles ($p < 0.05$ for all differences). Lastly, there was no statistically significant difference between the teacher and the companion role ($p = 0.267$).

Regarding the negative emotions a significant difference was also found in each emotion for the four roles ($F(2,432) = 4.821$, $p = 0.023$) for interest, $F(2,432) = 5.908$, $p = 0.001$) for inspiration and $F(2,432) = 5.261$, $p = 0.002$ for confident). Specifically, there was no statistically significant difference between nervous and tired ($p = 0.797$), while both of them were higher than shy ($p < 0.05$ for both). These differences were the same across all roles. The role associated with the highest negative emotions was the interviewer, followed by the teacher. However, no significant difference was found between them ($p = 0.119$). In total, the role of the interviewer presented significantly higher values than both customer guide and companion ($p < 0.05$ for both). The teacher role was not statistically significantly different from the companion role ($p = 0.075$), but was significantly higher than the customer guide ($p = 0.032$). Lastly, customer guide and companion roles did not differ ($p = 0.71$).

Summing up, positive emotions were presented at higher levels than negative ones. In general, participants felt more interested and confident than they felt inspired. Based on the questionnaire, the role of the interviewer concentrated the higher positive level and the teacher the highest negative one. However, the overall intensity of negative emotions was very low. Shyness was the less presented emotion whereas nervousness and tiredness were the most common ones. In regard to these emotions, interviewer was again the role with the highest levels, followed by the teacher, and lastly companion and customer guide. Figure 5.28 summarizes both positive and negative emotions for the four roles of Nadine.



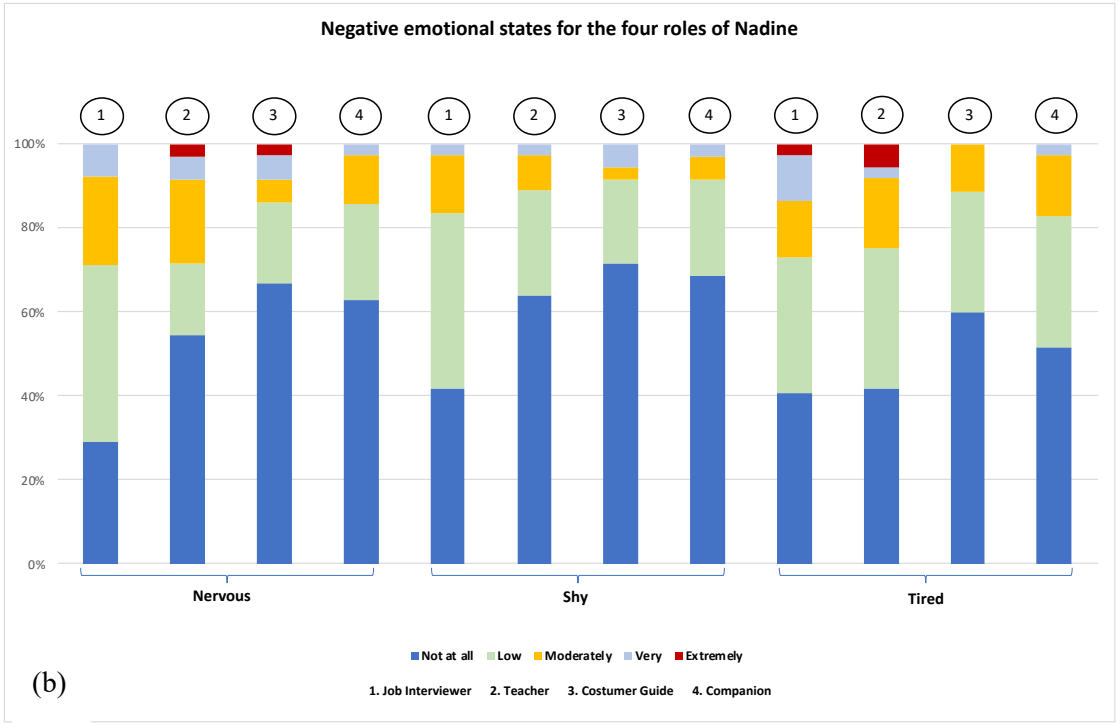


Figure 5.28 Positive (a) and Negative (b) emotional states as described by our participants for the four roles of Nadine.

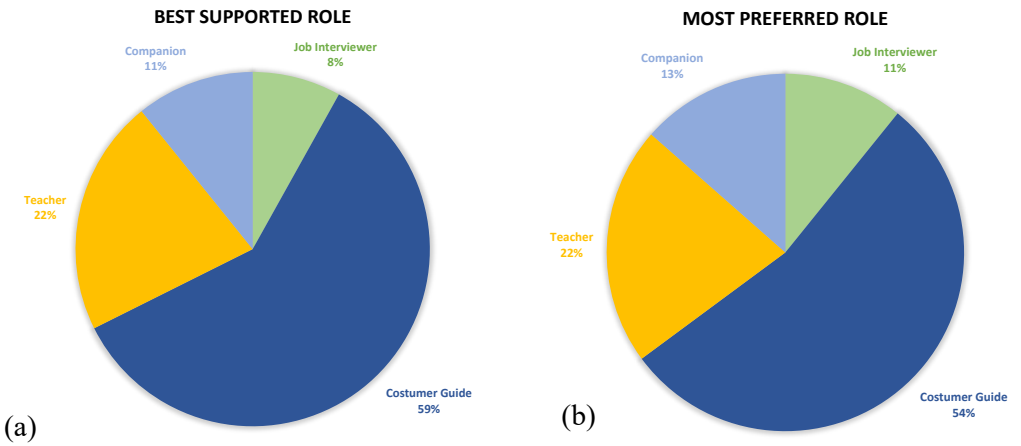


Figure 5.29 Results from our questionnaire regarding the preference of the four roles. *(a)*: the percentages for the best supported role. *(b)*: the percentages of the most preferred role according to the 40 participants.

Figure 5.29 presents the preference of the participants, as they declared it in the questionnaire. The interest of participants was more triggered by the role of companion, but they found that a robot is more friendly as a customer guide. We tend to believe that this is a result of the immediacy and the purity of this role, as everything is very precise and consequently quick. However, the role of the teacher was voted as the more sociable one.

5.3.3.2 Overall attitude towards robots

The NARS questionnaire is commonly used to examine the general attitude of participants towards robots. Our participants answered the questions on a Likert-type scale ranging from 1 (not at all) to 5 (extremely). The answers were coded, and averages were calculated. Three of the items were reverse coded to fit with the general direction of the answers (higher scores indicating more negative attitudes). Table 5.18 presents the questions used as well as the average scores. The average attitude towards robots was found relatively neutral (2.3). The participants gave higher scores for questions that referred to the idea of robots' dominance, e.g robots dominating humans in the future, being developed in human beings, or humans relying on robots. People also showed anxiety towards a possible need to use robots in their jobs. On the contrary, people expressed lower negative emotions when it came to standing in front of a robot or talking to one. They showed positive attitudes about being relaxed while talking to robots or having emotional connections with them.

Table 5.18 Average scores for the NARS questionnaire. Questions marked with a * are inverted.

Question	Score
I would feel uneasy if I was given a job where I had to use robots	2.52
I would feel nervous operating a robot in front of other people	2.17
The word "robot" means nothing to me	2.39
I would hate the idea that robots or artificial intelligence were making judgements about things	2.52
I would feel very nervous just standing in front of a robot	1.91
I would feel paranoid talking with a robot	2.04
I would feel uneasy if robots really had emotions	2.48
Something bad might happen if robots developed into living beings	2.74
I feel that if I depend on robots too much, something bad might happen	2.48
I am concerned that robots would be a bad influence on children	2.22
I feel that in the future society will be dominated by robots	2.57
I would feel relaxed talking with robots*	2.04
If robots had emotions, I would be able to make friends with them*	1.96
I feel comforted being with robots that have emotions*	2.09
Total	2.3

5.3.3 Discussion

In this study, participants were urged to interact with the humanoid social robot Nadine under four predefined roles: interviewer, teacher, customer guide, and companion. Behavioral and emotional reactions based on brain activity, body movements, and questionnaires were examined. Moreover, for a more complete assessment of our results, participants' general attitudes towards robots was also measured through a NARS questionnaire.

Firstly, the NARS results indicated that the participants had a relatively neutral attitude towards robots before the interaction. Their concern was mostly on robots' development and possible dominance over people in the future. However, they approached positively potential social or emotional interactions with robots.

The questionnaire revealed the emotional states of the participants throughout the procedure. Participants seemed to be most emotional about the role of the interviewer. The ambivalence of having high levels of both positive and negative emotions can be attributed to the fact that this role was the first to be presented but also to the stressful nature of an interview process by definition. Results could stem from the excitement and the surprise of the first experience. On the other hand, the teacher role seemed to be the least enjoyable

as participants reported the lowest level of positive emotions and high levels of negative ones. That indicated that such a context should be revised and improved. Customer guide and companion presented the same and lowest level of negative emotions but the former had higher positive ones. Thus, it seems that customer guide was the most enjoyable role, probably due to the usefulness and the practicality of the interaction, which is in line with previous studies indicating a willingness of people to interact with robots in that context [99, 138, 140]. The companion mode seemed to have a more neutral effect as participants experienced no unpleasant emotions but no positive either.

The recording of brain activity and body movements validated the above. EEG data were collected from five brain areas: prefrontal, frontal, parietal, temporal, and occipital. For all the areas except for the parietal, a pattern of increasing frequencies throughout the experience was noticed, indicating an effect of familiarity. In the prefrontal area, the interviewer role elicited a theta rhythm, which has been linked to spatial working memory [256]. Therefore, we can assume that participants engaged their working memory during the interview phase to retrieve the proper information to answer the questions. The alpha rhythm presented in the other roles shows a strong engagement of the decision making and behavior mechanisms.

Regarding the frontal area, theta oscillations were detected for the role of the interviewer which indicated that the task was demanding [230] and participants had to put a higher cognitive effort to get focused and interact with Nadine. Other studies have also noticed frontal theta oscillation during a first human-robot interaction [201]. Subsequently, for the second role of the teacher, alpha rhythm was detected and this rhythm can be associated with the participants' concentration. Frontal alpha rhythm is related to the origin of the top-down perceptual process and thus it reveals that the brain constructs the perception based on an existing, already registered, experience [258]. For the remaining two roles, a beta state was observed, which suggests that participants were not stressed but mentally active.

Regarding the parietal area, alpha rhythm was noticed for the roles of the interviewer and companion. This finding is in line with all our previous studies indicating that participants were unintentionally focused on stimuli they are not familiar with. As we mentioned before, alpha rhythm in the parietal cortex is also associated with emotional engagement and this explains its presence in this area for the role of the companion. For the roles of the teacher and the customer guide, a beta state was detected.

In the temporal lobe, a beta rhythm with an increase in the frequency for each role was detected. The temporal lobe is associated with auditory processing [230] and the beta rhythm shows that participants had no difficulties in following Nadine's speech. However, the role of the interviewer showed the lowest value of the beta state. This means that the participants faced the biggest difficulty in deciphering robot's voice while being interviewed, which may be explained by the fact that this role was the first one, and participants

had no prior robot experience. The more they interacted with the robot, the more familiar they became with it, in terms of audio processing [259].

Lastly, the results in the occipital lobe verified our previous findings regarding human-robot interaction. The interaction with the humanoid robot elicited alpha oscillations in this area (low alpha state for the role of interviewer).

Body movement results can be put up against the questionnaire and EEG activity, to verify the emotion of the participants in regard to the different roles. First, during the role of interviewer high levels of fear, anger, and sadness were presented amongst the participants. This is in line with the high negative emotions extracted by the questionnaire and the frontal theta oscillations found in brain activity, indicating that this role was unpleasant for the participants. Kinect data also showed relatively low levels of happiness and surprise, indicating that the questionnaire's positive emotions can be attributed to the excitement about the initial interaction with the robot. Moreover, the negative assessment of the teacher role based on the questionnaire was confirmed by the Kinect results. They indicated very low levels of happiness and surprise, along with very high levels of anger. Furthermore, the customer guide role was confirmed as eliciting high levels of positive emotions, such as happiness and surprise, along with, unexpectedly, high levels of anger. Finally, the role of companion actually showed the highest levels of positive emotions on the Kinect data, and an absence of negative emotions. Thus, we can assume that the companion role elicited positive emotions, that either they couldn't be detected by the questionnaire or the participants didn't admit them.

Summing up, the roles can be summarized as follows:

- The interviewer role was the one that elicited the most intensive reactions, both in terms of emotions and neural activity. This can be partially explained by the fact that his role was presented first and participants were unfamiliar with Nadine. The low alpha state in the occipital area can verify the latter. The questionnaire revealed high levels of interest and inspiration. Aside from this, the role seemed to be dominantly unpleasant for the participants and required a lot of cognitive activity, as in every normal job interview procedure. EEG results were in line with that, as they verified the emotional engagement with the existence of the alpha state in the parietal lobe. Moreover, prefrontal and frontal theta oscillations were dominant only for this role. However, participants voted that Nadine had the friendliest reactions under this role.
- The teacher role was the one with the most negative reactions. Kinect data and questionnaire revealed only negative emotions and tiredness. EEG data though showed a higher level of concentration as alpha waves in the frontal area appeared only for this role. Our results suggest that this role needs revision and improvement.

- The customer guide role was the one that elicited the best emotional response, in terms of the most positive and the least negative emotions. It did not require heavy cognitive activity, the questionnaire revealed a high confidence level, and participants showed no signs of stress as a low beta state was dominant in the frontal area. The customer guide was voted by the participants as the most preferred and suitable role for Nadine.
- Lastly, the companion role had mixed results. Body movements showed positive emotions that were not supported or captured by the questionnaire. Kinect data showed high levels of surprise and happiness though. EEG verified the latter, as this role engaged more alpha activity in the parietal lobe, associated with perception and emotional engagement. Moreover, a low beta state in the frontal area confirms the absence of stress and anxiety.

5.4 General Discussion

In this study, we conducted a multi-disciplinary, in-depth analysis of human responses derived from different kinds of social interactions: human-human, human-robot, and human-virtual human. Our research goal was bifold: to explore and compare human physiological, behavioral and emotional states and to examine the potential of social robots to execute different roles. We found that the human brain, body, and vocal responses changes based on the nature of the interlocutor and emotions are affected too. Different roles can also influence human acceptance towards robots, leading us to believe that nonhuman agents cannot penetrate all the domains of everyday human life.

Having gathered the results from both parts of our experiments, we can answer our fourth and last research question: *Can changes in agents' up-to-date technology and design be related to smoother, more pleasant, and efficient human-nonhuman interactions?* Based on our findings and the statements of the up-to-date literature, we do believe that the design of nonhuman agents and the technology behind them can affect human perception and behavior. The choice of an agent should always fit the general context of the interaction as apparently the acceptance and the preference are directly affected by the context and the environment.

Except their emotional states, our participants were asked to evaluate the performance of the agents. Not surprisingly, the human agent was voted as the most sociable and friendly, but Nadine followed. With all our physiological results, we expected to find Nicole in the second position. However, participants stated that the lack of eye gaze in the digital human played a crucial role, as it has already been mentioned in the literature [62]. Another limiting factor is the physical presence. No matter the negative emotions found, participants seemed to be more positive and motivated to interact with Nadine, which means that social

robots are very promising in a context of social interaction. Body reactions showed nervousness but also a bigger range of movement and calmer voice when interacting with Nadine, compared to Nicole. Thus, when interacting with a social robot, human behavior resembles more the one from H-H interaction. However, depending on the context of the interaction, digital humans may provide humans with a comfort that a robot cannot. The latter needs to be considered before designing or choosing a nonhuman agent. In general, humans surely need more time to become more familiar with this technology, but the latter also needs to meet more the human needs. Our participants reported that they had to change the content of their responses, according to the nature of their interlocutor, as they felt that the communication could not be equal. This was verified by the fact that emotions also were changed according to the nature of the agent.

Can different roles of a robot change the perception and preference of the participants? As we clearly showed, a different role can change completely the perception and the preference of a human towards a robot. We noticed that multiple exposures to robot interactions affects human physiological and psychological responses. It is though sure that culture play also a role in this perception as external stimuli are different and humans are not receptive in the same way.

This work[1][2][3][4] contributes to the existing SoA with the following conclusions:

- The role of agents' behavior.

Nowadays, research and reviews show us that the need is focused on designing social agents in a more human-like way behaviorwise and not in terms of appearance. Our study complements the above, as we

---

[1]Baka E., Mishra N., Magnenat Thalmann N., Frantzidis C., Vleioras G. (2022) Human, Avatar or a Robot? A multimodal analylsis towards uncovering human's perception and reactions to social interactions (Submitted)

[2] Baka E., Mishra N., Magnenat-Thalmann N. (2022) "Social robots and Digital humans as Job Interviewers: A study of humans' reactions towards a more natural interaction", International Conference on Human Computer Interaction (HCII) 2022 (Submitted and Accepted)

[3] Mishra, N., Baka, E., & Thalmann, N. M. (2021). Exploring Potential and Acceptance of Socially Intelligent Robot. In Intelligent Scene Modeling and Human-Computer Interaction (pp. 259-282). Springer, Cham. https://doi.org/10.1007/978-3-030-71002-6_15

[4] Baka, E., & Thalmann, N. M. (2021). Human—Technology Interaction: The State-of-the-Art and the Lack of Naturalism. In *Intelligent Scene Modeling and Human-Computer Interaction* (pp. 221-239). Springer, Cham.

concluded that it is the reactions of the agents that trigger the different human responses and not the appearance alone, as both our nonhuman agents reach a high level of human likeness. This complements the outcome of several studies that have examined the importance of the appearance and humanization [5, 32, 50, 52, 55, 56, 152, 213, 219]. We verified that it is important to orient the research towards humans, extracting humans' features, revealing human needs, and integrating them into the technology. Thus, we believe that the up-to-date technology should be more oriented towards the integration of human behavioral features that will elicit better cognitive and emotional responses and ensure a higher level of engagement through a new more naturalistic communication channel. Lastly, our model laid the groundwork for building a system that can classify the human voice based on the nature of the interlocutor and can guide engineers and designers to more human-like robots.

- The role of mimicry.

Moreover, this work verified the role of mimicry during social interactions, especially when the interlocutor is physically present, as described also in [17, 45]. As we saw, participants tended to use more their body when interacting with the human agent and their vocal behavior was close enough to the one of the agent. The interaction with Nadine also affected their reactions and participants stated that the eye gaze and the direction of the body played a role, as stated also in [73], compared to the digital human who was virtually present. However, the context of our interactions didn't allow us to find clear signs of emotional mimicry. Mimicry up to now has been studied mainly through facial expressions, thus facial EMG, or through questionnaires. We provided a different perspective via a multidisciplinary approach recording the whole body.

- Social signals input

The multidisciplinary approach of our study allowed us to report several human behavioral and physiological patterns during H-H and H-NH interactions, which weren't part of the literature yet. Audio data gave us different patterns of vocal behavior among the three interactions for the majority of the features measured. This was verified by our ML classification model that proved that the human voice clearly changes based on the nature of the interlocutor (human, robot, or avatar). EMG and Kinect data gave us significant information regarding the body reactions and their involvement in a social context. Moreover, physiological brain patterns, that were not reported before, were registered via our work.

CHAPTER 6

CONCLUSION

"Will robots inherit the earth? Yes, but they will be our children"

- *Marvin Minsky, American cognitive and computer scientist*

Conclusion

6.1 Discussion

Our work was concentrated on finding answers regarding human perception and behavior towards social interactions of different contexts, detecting and possibly covering the gap between human and nonhuman interactions. We examined technological environments, like Virtual reality and different kinds of nonhuman social agents, based on the up-to-date technology, aiming to delve deeper into the features and the nature of the natural H-H interaction and the continuously growing H-NH one. We used a multidisciplinary approach to cover limitations of previous studies and we explored how the design of an environment or an appearance and the physical presence of an agent can affect human perception, imagination, concentration, expectations, and emotions. We found common physiological and behavioral human patterns among all our studies that gave us insights on differences between humans and nonhumans agents as well as between the nonhuman themselves.

Regarding brain activity, we noticed frontal theta oscillations during any nonhuman interaction and exposure to a VE. Based on that, we concluded that people tend to unintentionally be more concentrated on an environment or an interaction they are not familiar with. However, we noticed that this can change after multiple exposures to a social robot. Similarly, alpha state was dominant in the parietal area for all interactions, leading us to the assumption that parietal alpha can act as an indicator of a technological-based interaction. We also found a dominant alpha state in the occipital area for all interactions, indicating the level of engagement of visual attention mechanisms. Moreover, temporal alpha oscillations were present during HRI and frontal delta oscillations during both H-H and HR Interactions. The latter appears due to the physical presence of the robot and the human, compared to the virtual human.

Regarding behavioral patterns, differences were obvious in voice and body movements. When interacting with a nonhuman agent, we tend to give shorter answers with longer pauses and to speak slower and louder. Frequency is higher, perturbations in amplitude are less and the value of HNR is bigger, showing that there is less hoarseness in the voice. The movements of the arm are smaller and sharper related to anger, there are more frequent head movements and shoulder elevations revealing discomfort, fingers present long and intense movements related to fear and the overall range of movement is narrower. Moreover, the differences in motion between the two body parts are bigger.

Emotional states between interactions were also different. In general, during H-H interaction participants declared happier, more confident but also shyer whereas during H-NH they felt mainly agitated and

surprised. It has already been stated that anxiety is the most common emotion in HRI, combined with a general emotional constraint [21, 32, 50, 218]. Voice and motion showed even sadness and fear. Brain activity revealed that participants tended to deal more with their emotions during H-NH interactions, and especially during H-NA. Moreover, it showed us that they were more concentrated, dealing most of the time with a higher cognitive load. This was also verified by the correlation of the upper body with the alpha band in several brain areas, like the Frontal, Temporal and Occipital, which is linked with the top-down procedure meaning with the transition from perception to action. However, a preference over the social robot was noted, verified also by a bigger range of motion during this interaction related to the emotion of happiness and an emotional arousal found by brain activity. Inspiration and interest were apparent while interacting with the social robot whereas audio data indicated a feeling of calm during interaction with the virtual human. That led us to conclude that the choice of a nonhuman agent should depend on the need of the tasks and the targeted users.

However, we noticed that the differences in humans' reactions between the interactions were obvious, as verified also by our classification model, showing that there is still a need for improvement to reach a level of desired human comfort. We verified the role of eye gaze and facial expressions already mentioned in the literature [62] but we noticed that the physical presence of an agent is also crucial for its acceptance. We confirmed brain patterns already discovered regarding VR, like the alpha state in the parietal area [194], and regarding HRI, like frontal theta oscillations [201] but we also discovered new patterns that haven't been mentioned before. It has already been stated that people tend to spend more time in the conversation with another human [17, 18] and we verified that during H-NH participants gave shorter answers and were less talkative. Moreover, we found a lot of differences between the human and the nonhuman vocal reactions, which indicates that further work should be done to better humanize the voice of nonhuman agents. Some features like response time and speechrate, are positively correlated which means that the more humanized a voice reaction of a nonhuman agent will be, the more natural the humans will speak. However, here we are wondering to what extent a social agent or a robot need to be human-like to fulfill the humans' need and expectations and to be socially accepted. It is worth mentioning that, although negative emotions were found during H-NH interactions, interest, inspiration, and motivation were also apparent and 72% of our participants in our last study confirmed feeling less nervous while discussing with the social robot or the virtual human. Thus, we believe that humans like the, unfamiliar yet, interaction with the nonhuman agents but to feel physically and emotionally comfortable with them a lot of work is still required. There are also studies proving that nonhuman agents can be more successful and efficient compared to humans under specific tasks [75, 76] However, the potential is big and promising. To complement the latter, we also verified the finding of Urgen et al. that the human brain cannot differentiate actions between robots and humans [201], showing that the difference is only at a visual perception level.

Lastly, regarding the VR environment, we concluded that the experience of the users is related to the level of presence they can feel. Moreover, presence is not related to the interaction into the VE but to the VR itself. The design of the VE, however, plays a role in the concentration of users, as realistic environments make the latter more conscious of the external environment. Thus, we believe that VR, for the time being, and under the existing technology, has the potential to become a useful, efficient, and reliable tool that can more adequately host a social agent.

6.2 Contributions

Facing the challenges for humanizing social agents and deciphering human complex social behavior, we proposed a multidisciplinary, in-depth analysis of human physiological, behavioral, and emotional reactions in different technological nonhuman interactions. We first examined the role of VR in human perception, indicating its efficacy in an educational context. We then continued with the study of human perception in human-humanoid interaction, examining its effects on cognitive and emotional states. We concluded with a complete, thorough study of humans' reactions in human-robot, human-avatar and human-human interaction, trying to shed some light on the thoughts and concerns of the use but also the humanization of social nonhuman agents. The study is accompanied by a thorough literature review and the extracted data have been analysed in detail and validated with statistical analysis as well as have been used for modeling aspects of human behavior.

This work contributes to the broad domain of Human – Robot Interaction, giving insights into fields like computer science, neuroscience, and psychology. We can mention four main contributions :

- *A new dataset of human reactions.*

First of all, our work provides a novel multimodal dataset of human reactions extracted from H-H and H-NH interactions, including participants of different ages, gender, and ethnicities. To the best of our knowledge, no such dataset has been recorded before, including multidisciplinary human physiological and psychological information. This dataset complements the already used dataset mentioned in Table 2.2 of Chapter 2 and can be used as a guide for HCI and HRI future research. The audio data that have already been modeled via ML techniques can directly be used for further research as emotion recognition.

- *Simultaneous recordings of physiological and psychological human data.*

One of our study's innovations is that it brings together a range of methods (EEG, EMG, motion capture, audio analysis, and psychometrics) that have been commonly used in HCI and HRI but in isolation. Thus, it allows us to direct analyse, compare and correlate and provide a comprehensive picture of the brain and

behavior of an individual who interacts with a human and nonhuman agent. We introduce a novel approach to study human-computer and human-human interaction and our work paves the way for future studies in other domains. New findings regarding brain activity and human behavioral patterns during nonhuman interactions can be added to the literature.

- *A voice model that can categorize the human voice based on the nature of the interlocutor.*

We provide a human voice model, derived from our audio data classification, able to differentiate the human voice based on the nature of the social interaction. Our results can potentially contribute to the design of nonhuman agents, as the direct comparison between humans' reactions and between humans and nonhumans made the needs clearer.

- *A transparent view of the robots' potential and acceptance.*

Lastly, the NARS questionnaire along with the robot testing under different roles provide a clear picture of human perception towards robots and replies to Figure 2.4. We complement and confirm aspects of the table regarding psychological and physical parameters. Moreover, we practically examined the domain of robotic psychology, theoretically presented by Stock et al. [21], verifying that human behavioral patterns are clearly differentiated between H-H and H-NH interaction. We also noticed that a lot of patterns are also different between human-avatar and human-robot interactions.

Overall, this work complement existing reviews examining the human perception towards robots or virtual humans as well as H-H interaction [47, 50]. Our system can be used as a base for further research to expand more the human side of the HRI and to support the social acceptance of nonhuman agents.

6.3 Limitations and future research

Despite the multidisciplinarity and the novel approach of our system, our work presents some limitations. First of all, the technology and the nonhuman agents used are based on the up-to-date state-of-the-art that continuously changes as the humans' needs and expectations advance. Thus, new research is constantly required to support this ever-increasing evolution. Moreover, given the nature of our study (physiological recordings, like EEG), it was difficult to find motivated female participants to contribute. Thus, our sample was not always balanced. Given also that most of the experiments took place in Singapore our ethnicity variable was not always balanced. To define how human's reactions are affected by culture or gender, a more diverse sample is required in future research.

Our model differentiates human reactions based on the nature of the interlocutor into the interaction but it is based only on the human voice. Moreover, due to the irregularities caused by the pandemic COVID-19,

we didn't have the opportunity to validate this model in a new sample of participants. Further research is required to expand this kind of model to more modalities, like movements or brain activity, that could clarify the patterns of human behavior and contribute better to the future design of nonhuman agents.

Lastly, our VR environment was based on an educational context and the concept of the last experiment was based on a job interview. We assume that humans' reactions have been affected by the specific scenarios and it would be interesting to expand this research in different contexts, allowing broader documentation of human responses.

APPENDIX A

PRELIMINARY RESEARCH PERFORMED FOR OTHER EUROPEAN RESEARCH PROJECTS

The work of this thesis was supported by several European projects. Albeit not completely relevant, each of these projects broadened the topic of this research, made the work more varied and diverse, and ensured further knowledge and techniques. Each section presents a brief description of the project and includes a section with the contribution related to the work presented in this thesis.

A1. NOTRE: Preliminary work on EEG and HR recordings and analysis

NOTRE project was a Horizon 2020 Twinning Programme aiming to develop a network that will strengthen and enhance the research and innovation potential of the newly established Social Computing Research Center (SCRC) at the Cyprus University of Technology. The ultimate goal was the stimulation of scientific excellence and innovation capacity in the area of Social Computing. Towards this goal, we worked closely with the GET Lab at the Cyprus University of Technology and created a VR platform for training purposes in the educational sector. Specifically, a VE, presenting students in a school under drug use, offering the experience from different perspectives and simulating the effects of possible drug use, was used to examine the overall user experience and the emotional states as well as the level of presence achieved. Figure 8.1 shows the different VEs used.

For the evaluation of the user's emotional and physiological situation a wireless EEG device EMOTIV Epoc+ was used, combined with a smartwatch for the recording of the HR and a questionnaire. For the VR, an Oculus Rift device was used. The detailed methodology of the experiment is described in [260].

For the analysis of the brain signal, EEGLab was used, a MATLAB toolbox. Figure 8.2 shows an example of a dominant frequency of two perspectives. Statistical analysis of EEG and HR was conducted in SPSS with Krusk-Wallis and Mann-Whitney tests.

Figure 7.1 The VE as seen from the different perspectives. Top: Teacher's perspective, Middle: Student drug user's perspective, Bottom: Healthy students' character perspective

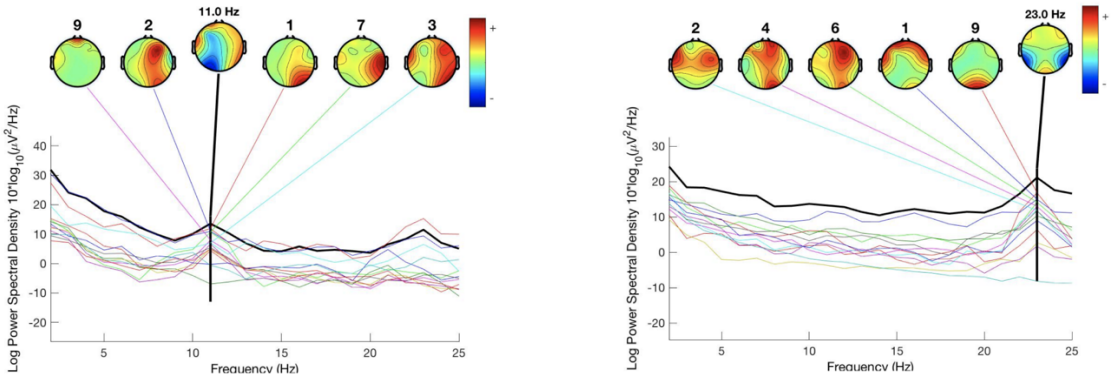


Figure 7.2 Dominant brain frequency for healthy student (Left) and teacher (Right) perspectives. The diagram was constructed after ICA was applied.

A1.1 Conclusion and Contribution

Our involvement in this project provided us with significant knowledge about:

- Design and conduction of experiments
- EEG and HR analysis
- The efficiency of VR systems for training purposes

A2. MINGEI: Face and Body reconstruction

MINGEI is a European Horizon 2020 project aiming at the representation and the preservation of tangible and intangible aspects of cultural Heritage Crafts. Specifically, there are three pilots, namely glass, mastic, and silk weaving, targeting to represent the knowledge that needs to be transmitted from master to apprentice. Towards this direction, during this projects partners work to capture the motion and the tool usage of HC practitioners and collect significant information from the Living Human Treasures and archive documentaries to preserve and illustrate skill and tool manipulation. The representation will be done in VEs as well as via a mobile application and thus, the existence of VHs is crucial as guides and masters.

For that, MIRALab started working on face and body reconstruction that could give realistic avatars looking alike exactly to the real persons. Thus, an image-based full body and face 3D scanner was used as shown in Figure 8.3. The setup of the cameras was adapted to the needs of the task. We took images for both face and full body. From these images, we proceeded to face and body reconstruction, as shown in Figure 8.4, with the Agisoft Metashape software[261].

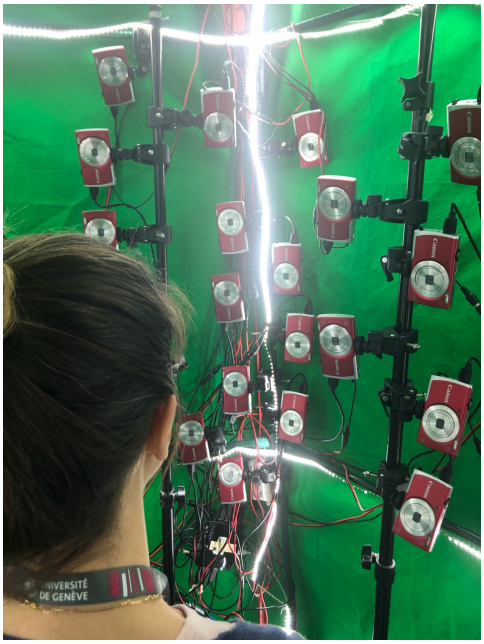


Figure 7.3 Image of the scanner in face position

Figure 7.4 Face and Body reconstruction

After the reconstruction, we adjusted the topology of the face to a standard head, as shown in figure 8.5 and then, we blended the textures and the morphology on one mesh. Then the blendshapes were generated. This procedure was made with the Visage Technologies software [262].

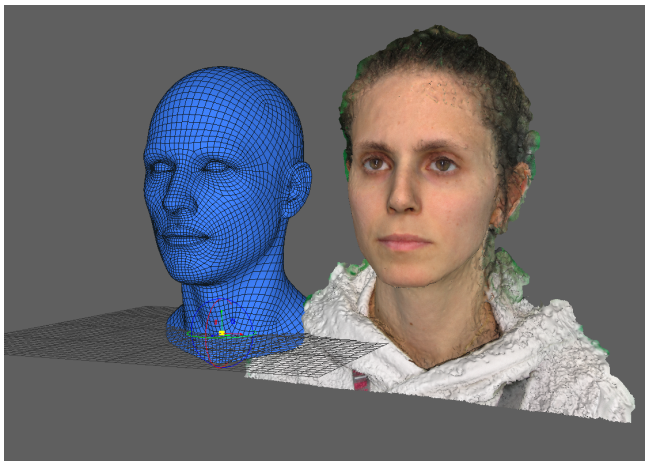


Figure 7.5 Face wrapping

A2.1 Conclusion and Contribution

Although this project is slightly irrelevant with the main research of this thesis, it gave us a strong insight into the reconstruction of realistic virtual humans.

BIBLIOGRAPHY

1. Ismail, WOAS Wan, M. Hanif, S. B. Mohamed, Noraini Hamzah, and Zairi Ismael Rizman. 2016. Human emotion detection via brain waves study by using electroencephalogram (EEG). *International Journal on Advanced Science, Engineering and Information Technology* 6: 1005–1011.
2. Ekman, Paul. 1973. Cross-cultural studies of facial expression. *Darwin and facial expression: A century of research in review* 169222.
3. Poria, Soujanya, Erik Cambria, Rajiv Bajpai, and Amir Hussain. 2017. A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion* 37. Elsevier: 98–125.
4. Nass, Clifford, and Youngme Moon. 2000. Machines and mindlessness: Social responses to computers. *Journal of social issues* 56. Blackwell Publishers Inc. Boston, USA and Oxford, UK: 81–103.
5. Baylor, Amy L. 2011. The design of motivational agents and avatars. *Educational Technology Research and Development* 59. Springer: 291–300.
6. Ferber, Jacques, and Gerhard Weiss. 1999. *Multi-agent systems: an introduction to distributed artificial intelligence*. Vol. 1. Addison-Wesley Reading.
7. Künecke, Janina, Andrea Hildebrandt, Guillermo Recio, Werner Sommer, and Oliver Wilhelm. 2014. Facial EMG responses to emotional expressions are related to emotion perception ability. *PloS one* 9. Public Library of Science San Francisco, USA: e84053.
8. Enea, Violeta, and Sorina Iancu. 2016. Processing emotional body expressions: state-of-the-art. *Social neuroscience* 11. Taylor & Francis: 495–506.
9. Schrammel, Franziska, Sebastian Pannasch, Sven-Thomas Graupner, Andreas Mojzisch, and Boris M. Velichkovsky. 2009. Virtual friend or threat? The effects of facial expression and gaze interaction on psychophysiological responses and emotional experience. *Psychophysiology* 46. Wiley Online Library: 922–931.
10. De Gelder, Beatrice. 2009. Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364. The Royal Society: 3475–3484.
11. Kret, Mariska E., and Beatrice de Gelder. 2013. When a smile becomes a fist: the perception of facial and bodily expressions of emotion in violent offenders. *Experimental Brain Research* 228. Springer: 399–410.
12. Spence, Patric R. 2019. *Searching for questions, original thoughts, or advancing theory: Human-machine communication*. Elsevier.
13. Inoue, Koji, Kohei Hara, Divesh Lala, Shizuka Nakamura, Katsuya Takanashi, and Tatsuya Kawahara. 2021. A job interview dialogue system with autonomous android ERICA. In *Increasing Naturalness and Flexibility in Spoken Dialogue Interaction: 10th International Workshop on Spoken Dialogue Systems*, 291–297. Springer Singapore.

14. Wairagkar, Maitreyee, Maria R. Lima, Daniel Bazo, Richard Craig, Hugo Weissbart, Appolinaire C. Etoundi, Tobias Reichenbach, Prashant Iyengar, Sneh Vaswani, and Christopher James. 2021. Emotive response to a hybrid-face robot and translation to consumer social robots. *IEEE Internet of Things Journal*. IEEE.

15. Tulsulkar, Gauri, Nidhi Mishra, Nadia Magnenat Thalmann, Hwee Er Lim, Mei Ping Lee, and Siok Khoong Cheng. 2021. Can a humanoid social robot stimulate the interactivity of cognitively impaired elderly? A thorough study based on computer vision methods. *The Visual Computer*. Springer: 1–20.

16. Li, Jamy, René Kizilcec, Jeremy Bailenson, and Wendy Ju. 2016. Social robots and virtual agents as lecturers for video instruction. *Computers in Human Behavior* 55. Elsevier: 1222–1230.

17. Mou, Yi, and Kun Xu. 2017. The media inequality: Comparing the initial human-human and human-AI social interactions. *Computers in Human Behavior* 72. Elsevier: 432–440.

18. Shechtman, Nicole, and Leonard M. Horowitz. 2003. Media inequality in conversation: how people behave differently when interacting with computers and people. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 281–288.

19. Fischer, Kerstin, Kilian Foth, Katharina J. Rohlfing, and Britta Wrede. 2011. Mindful tutors: Linguistic choice and action demonstration in speech to infants and a simulated robot. *Interaction Studies* 12. John Benjamins: 134–161.

20. Zawieska, K., M. Ben Moussa, Brian R. Duffy, and Nadia Magnenat-Thalmann. 2012. The role of imagination in Human-Robot Interaction. In *25th Annual Conference on Computer Animation and Social Agents, Autonomous Social Robots and Virtual Humans Workshop*.

21. Stock, Ruth, and Mai Anh Nguyen. 2019. Robotic psychology. What do we know about human-robot interaction and what do we still need to learn? In *Proceedings of the 52nd Hawaii international conference on system sciences*.

22. Kasap, Zerrin, and Nadia Magnenat-Thalmann. 2012. Building long-term relationships with virtual and robotic characters: the role of remembering. *The Visual Computer* 28. Springer: 87–97.

23. Kasap, Zerrin, Maher Ben Moussa, Parag Chaudhuri, and Nadia Magnenat-Thalmann. 2009. Making them remember—Emotional virtual characters with memory. *IEEE Computer Graphics and Applications* 29. IEEE: 20–29.

24. Zhang, Juzheng, Nadia Magnenat Thalmann, and Jianmin Zheng. 2016. Combining memory and emotion with dialog on social companion: A review. In *Proceedings of the 29th international conference on computer animation and social agents*, 1–9.

25. Moussa, Maher Ben, and Nadia Magnenat-Thalmann. 2013. Toward socially responsible agents: integrating attachment and learning in emotional decision-making. *Computer Animation and Virtual Worlds* 24. Wiley Online Library: 327–334.

26. Tahir, Yasir, Debsubhra Chakraborty, Tomasz Maszczyk, Shoko Dauwels, Justin Dauwels, Nadia Thalmann, and Daniel Thalmann. 2015. Real-time sociometrics from audio-visual features for two-person dialogs. In *2015 IEEE International Conference on Digital Signal Processing (DSP)*, 823–827. IEEE.

27. Vinciarelli, Alessandro, Anna Esposito, Elisabeth André, Francesca Bonin, Mohamed Chetouani, Jeffrey F. Cohn, Marco Cristani, Ferdinand Fuhrmann, Elmer Gilmartin, and Zakia Hammal. 2015. Open challenges in modelling, analysis and synthesis of human behaviour in human–human and human–machine interactions. *Cognitive Computation* 7. Springer: 397–413.

28. Esposito, Anna, Antonietta M. Esposito, and Carl Vogel. 2015. Needs and challenges in human computer interaction for processing social emotional information. *Pattern Recognition Letters* 66. Elsevier: 41–51.

29. Perez-Gaspar, Luis-Alberto, Santiago-Omar Caballero-Morales, and Felipe Trujillo-Romero. 2016. Multimodal emotion recognition with evolutionary computation for human-robot interaction. *Expert Systems with Applications* 66. Elsevier: 42–61.

30. Chanel, Guillaume, and Christian Mühl. 2015. Connecting brains and bodies: applying physiological computing to support social interaction. *Interacting with Computers* 27. OUP: 534–550.

31. Torres-Valencia, Cristian A., Hernan F. Garcia-Arias, Mauricio A. Alvarez Lopez, and Alvaro A. Orozco-Gutiérrez. 2014. Comparative analysis of physiological signals and electroencephalogram (EEG) for multimodal emotion recognition using generative models. In *2014 XIX Symposium on Image, Signal Processing and Artificial Vision*, 1–5. IEEE.

32. Łupkowski, Paweł, and Marta Gierszewska. 2019. Attitude towards humanoid robots and the uncanny valley hypothesis. *Foundations of Computing and Decision Sciences* 44: 101–119.

33. Dasgupta, Poorna Banerjee. 2017. Detection and analysis of human emotions through voice and speech pattern processing. *arXiv preprint arXiv:1710.10198*.

34. Johnstone, Tom. 2017. The effect of emotion on voice production and speech acoustics. Thesis Commons.

35. Melzer, Ayelet, Tal Shafir, and Rachelle Palnick Tsachor. 2019. How do we recognize emotion from movement? Specific motor components contribute to the recognition of each emotion. *Frontiers in psychology* 10. Frontiers: 1389.

36. Rimmele, Johanna M., Joachim Gross, Sophie Molholm, and Anne Keitel. 2018. Brain oscillations in human communication. *Frontiers in Human Neuroscience* 12. Frontiers: 39.

37. Aftanas, L. I., A. A. Varlamov, S. V. Pavlov, V. P. Makhnev, and N. V. Reva. 2001. Affective picture processing: event-related synchronization within individually defined human theta band is modulated by valence dimension. *Neuroscience letters* 303. Elsevier: 115–118.

38. Keil, Andreas, Matthias M. Müller, Thomas Gruber, Christian Wienbruch, Margarita Stolarova, and Thomas Elbert. 2001. Effects of emotional arousal in the cerebral hemispheres: a study of oscillatory brain activity and event-related potentials. *Clinical neurophysiology* 112. Elsevier: 2057–2068.

39. Klados, Manousos A., Christos Frantzidis, Ana B. Vivas, Christos Papadelis, Chrysa Lithari, Costas Pappas, and Panagiotis D. Bamidis. 2009. A framework combining delta event-related oscillations (EROs) and synchronisation effects (ERD/ERS) to study emotional processing. *Computational intelligence and neuroscience* 2009. Hindawi.

40. den Stock, Jan Van, Mathieu Vandenbulcke, Charlotte BA Sinke, and Beatrice de Gelder. 2014. Affective scenes influence fear perception of individual body expressions. *Human brain mapping* 35. Wiley Online Library: 492–502.

41. Baka, Evangelia, Mike Kentros, George Papagiannakis, and Nadia Magnenat-Thalmann. 2018. Virtual Reality Rehabilitation Based on Neurologic Music Therapy: A Qualitative Preliminary Clinical Study. In *International Conference on Learning and Collaboration Technologies*, 113–127. Springer.

42. Birmingham, Chris, Zijian Hu, Kartik Mahajan, Eli Reber, and Maja J. Matarić. 2020. Can I trust you? A user study of robot mediation of a support group. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 8019–8026. IEEE.

43. Šabanović, Selma, Casey C. Bennett, Wan-Ling Chang, and Lesa Huber. 2013. PARO robot affects diverse interaction modalities in group sensory therapy for older adults with dementia. In *2013 IEEE 13th international conference on rehabilitation robotics (ICORR)*, 1–6. IEEE.

44. Magnenat-Thalmann, Nadia, and Zhijun Zhang. 2014. Social robots and virtual humans as assistive tools for improving our quality of life. In *2014 5th International Conference on Digital Home*, 1–7. IEEE.

45. Hofree, Galit, Paul Ruvolo, Marian Stewart Bartlett, and Piotr Winkielman. 2014. Bridging the mechanical and the human mind: spontaneous mimicry of a physically present android. *PloS one* 9. Public Library of Science San Francisco, USA: e99934.

46. Mori, Masahiro, Karl F. MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine* 19. IEEE: 98–100.

47. Nowak, Kristine L., and Jesse Fox. 2018. Avatars and computer-mediated communication: a review of the definitions, uses, and effects of digital representations. *Review of Communication Research* 6. ESP: 30–53.

48. Stein, Jan-Philipp, and Peter Ohler. 2017. Venturing into the uncanny valley of mind—The influence of mind attribution on the acceptance of human-like characters in a virtual reality setting. *Cognition* 160. Elsevier: 43–50.

49. Edwards, Autumn, Chad Edwards, David Westerman, and Patric R. Spence. 2019. Initial expectations, interactions, and beyond with social robots. *Computers in Human Behavior* 90. Elsevier: 308–314.

50. de Borst, Aline W., and Beatrice de Gelder. 2015. Is it the real deal? Perception of virtual characters versus humans: an affective cognitive neuroscience perspective. *Frontiers in psychology* 6. Frontiers: 576.

51. Burleigh, Tyler J., Jordan R. Schoenherr, and Guy L. Lacroix. 2013. Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in human behavior* 29. Elsevier: 759–771.

52. Yamada, Yuki, Takahiro Kawabe, and Keiko Ihaya. 2013. Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. *Japanese psychological research* 55. Wiley Online Library: 20–32.

53. Rosenberg-Kima, Rinat B., E. Ashby Plant, Celeste E. Doerr, and Amy L. Baylor. 2010. The influence of computer-based model's race and gender on female students' attitudes and beliefs towards engineering. *Journal of Engineering Education* 99. Wiley Online Library: 35–44.

54. Shiban, Youssef, Iris Schelhorn, Verena Jobst, Alexander Hörnlein, Frank Puppe, Paul Pauli, and Andreas Mühlberger. 2015. The appearance effect: Influences of virtual agent features on performance and motivation. *Computers in Human Behavior* 49. Elsevier: 5–11.

55. Kätsyri, Jari, Klaus Förger, Meeri Mäkäräinen, and Tapio Takala. 2015. A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in psychology* 6. Frontiers: 390.

56. Ciechanowski, Leon, Aleksandra Przegalinska, Mikolaj Magnuski, and Peter Gloor. 2019. In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems* 92. Elsevier: 539–548.

57. Pfeifer, Rolf, and Christian Scheier. 2001. *Understanding intelligence*. MIT press.

58. Li, Jamy. 2015. The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies* 77. Elsevier: 23–37.

59. Milgram, Paul, Haruo Takemura, Akira Utsumi, and Fumio Kishino. 1995. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, 2351:282–292. International Society for Optics and Photonics.

60. Zhao, Shanyang. 2003. Toward a taxonomy of copresence. *Presence* 12. MIT Press: 445–455.

61. Lee, Kwan Min, Younbo Jung, Jaywoo Kim, and Sang Ryong Kim. 2006. Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. *International journal of human-computer studies* 64. Elsevier: 962–973.

62. Mollahosseini, Ali, Hojjat Abdollahi, Timothy D. Sweeny, Ron Cole, and Mohammad H. Mahoor. 2018. Role of embodiment and presence in human perception of robots' facial cues. *International Journal of Human-Computer Studies* 116. Elsevier: 25–39.

63. Slater, Mel. 2003. A note on presence terminology. *Presence connect* 3. Citeseer: 1–5.

64. Herbelin, Bruno, Roy Salomon, Andrea Serino, Olaf Blanke, Andrea Gaggioli, Alois Ferscha, Giuseppe Riva, Stephen Dunne, and Isabelle Viaud-Delmon. 2016. 5. Neural Mechanisms of Bodily Self-Consciousness and the Experience of Presence in Virtual Reality. In *Human Computer Confluence*, 80–96. De Gruyter Open Poland.

65. Hoffmann, Laura, and Nicole C. Krämer. 2011. How Should an Artificial Entity be Embodied? In *HRI 2011 Workshop*. Vol. 8.

66. Huang, Wei, Judith S. Olson, and Gary M. Olson. 2002. Camera angle affects dominance in video-mediated communication. In *CHI'02 Extended Abstracts on Human Factors in Computing Systems*, 716–717.

67. Shinozawa, Kazuhiko, Futoshi Naya, Junji Yamato, and Kiyoshi Kogure. 2005. Differences in effect of robot and screen agent recommendations on human decision-making. *International journal of human-computer studies* 62. Elsevier: 267–279.

68. Han, Shihui, Yi Jiang, Glyn W. Humphreys, Tiangang Zhou, and Peng Cai. 2005. Distinct neural substrates for the perception of real and virtual visual worlds. *NeuroImage* 24. Elsevier: 928–935.

69. Looije, Rosemarijn, Anna van der Zalm, Mark A. Neerincx, and Robbert-Jan Beun. 2012. Help, I need some body the effect of embodiment on playful learning. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, 718–724. IEEE.

70. Jost, Céline, Vanessa André, Brigitte Le Pévédic, Alban Lemasson, Martine Hausberger, and Dominique Duhaut. 2012. Ethological evaluation of Human-Robot Interaction: are children more efficient and motivated with computer, virtual agent or robots? In *2012 IEEE international conference on Robotics and Biomimetics (ROBIO)*, 1368–1373. IEEE.

71. Hasegawa, Dai, Justine Cassell, and Kenji Araki. 2010. The role of embodiment and perspective in direction-giving systems. In *2010 AAAI Fall Symposium Series*.

72. Moody, Eric J., Daniel N. McIntosh, Laura J. Mann, and Kimberly R. Weisser. 2007. More than mere mimicry? The influence of emotion on rapid facial reactions to faces. *Emotion* 7. American Psychological Association: 447.

73. Marschner, Linda, Sebastian Pannasch, Johannes Schulz, and Sven-Thomas Graupner. 2015. Social communication with virtual agents: The effects of body and gaze direction on attention and

emotional responding in human observers. *International Journal of Psychophysiology* 97. Elsevier: 85–92.

74. Kluttz, Nathan L., Brandon R. Mayes, Roger W. West, and Dave S. Kerby. 2009. The effect of head turn on the perception of gaze. *Vision research* 49. Elsevier: 1979–1993.

75. Yokotani, Kenji, Gen Takagi, and Kobun Wakashima. 2018. Advantages of virtual agents over clinical psychologists during comprehensive mental health interviews using a mixed methods design. *Computers in human behavior* 85. Elsevier: 135–145.

76. Lucas, Gale M., Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior* 37. Elsevier: 94–100.

77. Lazzeri, Nicole, Daniele Mazzei, Alberto Greco, Annalisa Rotesi, Antonio Lanatà, and Danilo Emilio De Rossi. 2015. Can a humanoid face be expressive? A psychophysiological investigation. *Frontiers in bioengineering and biotechnology* 3. Frontiers: 64.

78. Kompatsiari, Kyveli, Jairo Pérez-Osorio, Davide De Tommaso, Giorgio Metta, and Agnieszka Wykowska. 2018. Neuroscientifically-grounded research for improved human-robot interaction. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 3403–3408. IEEE.

79. Eyssel, Friederike, Laura De Ruiter, Dieta Kuchenbrandt, Simon Bobinger, and Frank Hegel. 2012. 'If you sound like me, you must be more human': On the interplay of robot and user features on human-robot acceptance and anthropomorphism. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 125–126. IEEE.

80. McGinn, Conor, and Ilaria Torre. 2019. Can you tell the robot by the voice? an exploratory study on the role of voice in the perception of robots. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 211–221. IEEE.

81. Mara, Martina, Simon Schreibelmayr, and Franz Berger. 2020. Hearing a nose? User expectations of robot appearance induced by different robot voices. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 355–356.

82. Goode, Lauren. 2018. How Google's eerie robot phone calls hint at AI's future. *Gear. Wired,. May 8th*.

83. Lubold, Nichola, Erin Walker, and Heather Pon-Barry. 2016. Effects of voice-adaptation and social dialogue on perceptions of a robotic learning companion. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 255–262. IEEE.

84. Lubold, Nichola, Erin Walker, Heather Pon-Barry, and Amy Ogan. 2018. Automated pitch convergence improves learning in a social, teachable robot for middle school mathematics. In *International conference on artificial intelligence in education*, 282–296. Springer.

85. Cambre, Julia, and Chinmay Kulkarni. 2019. One voice fits all? Social implications and research challenges of designing voices for smart devices. *Proceedings of the ACM on human-computer interaction* 3. ACM New York, NY, USA: 1–19.

86. Mayer, Richard E., and C. Scott DaPra. 2012. An embodiment effect in computer-based learning with animated pedagogical agents. *Journal of Experimental Psychology: Applied* 18. American Psychological Association: 239.

87. Kellermann, Kathy. 1992. Communication: Inherently strategic and primarily automatic. *Communications Monographs* 59. Taylor & Francis: 288–300.

88. Reeves, Byron, and Clifford Nass. 1996. *The media equation: How people treat computers, television, and new media like real people*. Cambridge university press Cambridge, United Kingdom.

89. Walther, Joseph B. 1996. Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication research* 23. Sage Publications London: 3–43.

90. Mischel, Walter, and Yuichi Shoda. 1995. A cognitive-affective system theory of personality: reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological review* 102. American Psychological Association: 246.

91. Mischel, Walter. 2004. Toward an integrative science of the person. *Annu. Rev. Psychol.* 55. Annual Reviews: 1–22.

92. Kalegina, Alisa, Grace Schroeder, Aidan Allchin, Keara Berlin, and Maya Cakmak. 2018. Characterizing the design space of rendered robot faces. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 96–104.

93. Antoniol, Giuliano, Roldano Cattoni, Mauro Cettolo, and Marcello Federico. 1993. Robust speech understanding for robot telecontrol. In *Proceedings of the 6th International Conference on Advanced robotics*, 205–209.

94. Burgard, Wolfram, Armin B. Cremers, Dieter Fox, Dirk Hähnel, Gerhard Lakemeyer, Dirk Schulz, Walter Steiner, and Sebastian Thrun. 1998. The interactive museum tour-guide robot. In *Aaai/iaai*, 11–18.

95. Mavridis, Nikolaos. 2015. A review of verbal and non-verbal human–robot interactive communication. *Robotics and Autonomous Systems* 63. Elsevier: 22–35.

96. Wada, Kazuyoshi, and Takanori Shibata. 2007. Living with seal robots—its sociopsychological and physiological influences on the elderly at a care house. *IEEE transactions on robotics* 23. IEEE: 972–980.

97. Dautenhahn, Kerstin, Michael Walters, Sarah Woods, Kheng Lee Koay, Chrystopher L. Nehaniv, A. Sisbot, Rachid Alami, and Thierry Siméon. 2006. How may I serve you? A robot companion approaching a seated person in a helping context. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, 172–179.

98. Makatchev, Maxim, Imran Fanaswala, Ameer Abdulsalam, Brett Browning, Wael Ghazzawi, Majd Sakr, and Reid Simmons. 2010. Dialogue patterns of an arabic robot receptionist. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 167–168. IEEE.

99. Kanda, Takayuki, Masahiro Shiomi, Zenta Miyashita, Hiroshi Ishiguro, and Norihiro Hagita. 2009. An affective guide robot in a shopping mall. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, 173–180.

100. Yamazaki, Akiko, Keiichi Yamazaki, Takaya Ohyama, Yoshinori Kobayashi, and Yoshinori Kuno. 2012. A techno-sociological solution for designing a museum guide robot: Regarding choosing an appropriate visitor. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 309–316. IEEE.

101. Petersen, Klaus, Jorge Solis, and Atsuo Takanishi. 2010. Musical-based interaction system for the Waseda Flutist Robot. *Autonomous Robots* 28. Springer: 471–488.

102. Kosuge, Kazuhiro, Tomohiro Hayashi, Yasuhisa Hirata, and Ryosuke Tobiyama. 2003. Dance partner robot-ms dancer. In *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)(Cat. No. 03CH37453)*, 4:3459–3464. IEEE.

103. Winograd, Terry. 1972. Understanding natural language. *Cognitive psychology* 3. Elsevier: 1–191.

104. Johnson, W. Lewis, Jeff W. Rickel, and James C. Lester. 2000. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial intelligence in education* 11. Citeseer: 47–78.

105. De Rosis, Fiorella, Catherine Pelachaud, Isabella Poggi, Valeria Carofiglio, and Berardina De Carolis. 2003. From Greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent. *International journal of human-computer studies* 59. Elsevier: 81–118.

106. Breazeal, Cynthia, and Juan Velásquez. 1998. Toward teaching a robot 'infant'using emotive communication acts. In . Citeseer.

107. Breazeal, Cynthia. 2003. Emotion and sociable humanoid robots. *International journal of human-computer studies* 59. Elsevier: 119–155.

108. Sugano, Shigeki, and Ichiro Kato. 1987. WABOT-2: Autonomous robot with dexterous finger-arm--Finger-arm coordination control in keyboard performance. In *Proceedings. 1987 IEEE International Conference on Robotics and Automation*, 4:90–97. IEEE.

109. Shibata, Takanori, Makoto Yoshida, and Junji Yamato. 1997. Artificial emotional creature for human-machine interaction. In *1997 IEEE international conference on systems, man, and cybernetics. Computational cybernetics and simulation*, 3:2269–2274. IEEE.

110. Graf, Birgit, Matthias Hans, and Rolf D. Schraft. 2004. Care-O-bot II—Development of a next generation robotic home assistant. *Autonomous robots* 16. Springer: 193–205.

111. Odashima, Tadashi, Masaki Onishi, Kenji Tahara, Kentaro Takagi, Fumihiko Asano, Yo Kato, Hiromichi Nakashima, Yuichi Kobayashi, Toshiharu Mukai, and Zhiwei Luo. 2006. A soft human-interactive robot ri-man. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1–1. IEEE.

112. Billard, Aude, Ben Robins, Jacqueline Nadel, and Kerstin Dautenhahn. 2007. Building Robota, a mini-humanoid robot for the rehabilitation of children with autism. *Assistive Technology* 19. Taylor & Francis: 37–49.

113. Patrizia, Marti, Moderini Claudio, Giusti Leonardo, and Pollini Alessandro. 2009. A robotic toy for children with special needs: From requirements to design. In *2009 IEEE International Conference on Rehabilitation Robotics*, 918–923. IEEE.

114. Dautenhahn, Kerstin, Chrystopher L. Nehaniv, Michael L. Walters, Ben Robins, Hatice Kose-Bagci, N. Assif, and Mike Blow. 2009. KASPAR–a minimally expressive humanoid robot for human–robot interaction research. *Applied Bionics and Biomechanics* 6. IOS Press: 369–397.

115. Hoffman, Guy, and Gil Weinberg. 2010. Shimon: an interactive improvisational robotic marimba player. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, 3097–3102.

116. Ramanathan, Manoj, Nidhi Mishra, and Nadia Magnenat Thalmann. 2019. Nadine humanoid social robotics platform. In *Computer Graphics International Conference*, 490–496. Springer.

117. Retto, Jesús. 2017. Sophia, first citizen robot of the world. *ResearchGate, URL: https://www. researchgate. net*.

118. Mišeikis, Justinas, Pietro Caroni, Patricia Duchamp, Alina Gasser, Rastislav Marko, Nelija Mišeikienė, Frederik Zwilling, Charles de Castelbajac, Lucas Eicher, and Michael Früh. 2020. Lio-a personal robot assistant for human-robot interaction and care applications. *IEEE Robotics and Automation Letters* 5. IEEE: 5339–5346.

119. Savage, Maddy. 2019. Meet Tengai, the job interview robot who won't judge you. *BBC Oline* 12.

120. QTrobot - expressive humanoid social robot for research and teaching. 2022. *LuxAI S.A.* https://luxai.com/humanoid-social-robot-for-research-and-teaching/. Accessed February 1.

121. Ameca. 2022. *Engineered Arts*. https://www.engineeredarts.co.uk/robot/ameca/. Accessed February 1.

122. Tsiourti, Christiana, Astrid Weiss, Katarzyna Wac, and Markus Vincze. 2019. Multimodal integration of emotional signals from voice, body, and context: Effects of (in) congruence on emotion recognition and attitudes towards robots. *International Journal of Social Robotics* 11. Springer: 555–573.

123. University, La Trobe. 2021. Research Centre for Computers, Communication and Social Innovation. https://www.latrobe.edu.au/reccsi. Accessed December 7.

124. Bepa. 2021. https://hr.robotvera.ru/static/newrobot/page384319.html. Accessed December 7.

125. Money, Ryan, Mark Newman, and Jacob Hanson. 2007. *On-line interview processing*. Google Patents.

126. AB, Anna & Hubert Labs. 2021. Hubert+1 - Add more to your team. https://hubert.ai/blog/. Accessed December 7.

127. Fridin, Marina. 2014. Storytelling by a kindergarten social assistive robot: A tool for constructive learning in preschool education. *Computers & education* 70. Elsevier: 53–64.

128. Fridin, Marina, and Mark Belokopytov. 2014. Robotics agent coacher for CP motor function (RAC CP Fun). *Robotica* 32. Cambridge University Press: 1265–1279.

129. Alemi, Minoo, Ali Meghdari, and Maryam Ghazisaedy. 2015. The impact of social robotics on L2 learners' anxiety and attitude in English vocabulary acquisition. *International Journal of Social Robotics* 7. Springer: 523–535.

130. Ivanov, Stanislav Hristov. 2016. Will robots substitute teachers? In *12th International Conference "Modern science, business and education*, 27–29.

131. Rossi, Piergiuseppe, Laura Fedeli, Silvia Biondi, Patrizia Magnoler, Anna Bramucci, and Cristiana Lancioni. 2015. The use of video recorded classes to develop teacher professionalism: the experimentation of a curriculum. *Journal of e-Learning and Knowledge Society* 11. Italian e-Learning Association.

132. Schneider, David R., and Clare Van Den Blink. 2006. An introduction to the nasa robotics alliance cadets program. *National defense education and innovation initiative report*.

133. Do, Ha M., Craig J. Mouser, Ye Gu, Weihua Sheng, Sam Honarvar, and Tingting Chen. 2013. An open platform telepresence robot with natural human interface. In *2013 IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems*, 81–86. IEEE.

134. Pandey, Amit Kumar, and Rodolphe Gelin. 2017. Humanoid robots in education: a short review. *Humanoid robotics: a reference*: 1–16.

135. Newton, Douglas P., and Lynn D. Newton. 2019. Humanoid robots as teachers and a proposed code of practice. In *Frontiers in education*, 4:125. Frontiers.

136. Lu, Vinh Nhat, Jochen Wirtz, Werner H. Kunz, Stefanie Paluch, Thorsten Gruber, Antje Martins, and Paul G. Patterson. 2020. Service robots, customers and service employees: what can we learn from the academic literature and where are the gaps? *Journal of Service Theory and Practice*. Emerald Publishing Limited.

137. O'brien, Matt. 2019. Will robots take your job? Quarter of US workers at risk. *AP News*.

138. Stock, Ruth Maria, and Moritz Merkle. 2017. A service Robot Acceptance Model: User acceptance of humanoid robots during service encounters. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 339–344. IEEE.

139. The social robot Nadine is working in the insurance company support desk since this October. 2021. https://lifeinsurance.kz/en/mirovoy-opyt/s-oktyabrya-etogo-goda-v-sluzhbe-podderzhki-strahovoy-kompanii-rabotaet-socialnyy-robot-nadin. Accessed December 7.

140. Heikkilä, Päivi, Hanna Lammi, Marketta Niemelä, Kathleen Belhassein, Guillaume Sarthou, Antti Tammela, Aurélie Clodic, and Rachid Alami. 2019. Should a robot guide like a human? A qualitative four-phase study of a shopping mall robot. In *International Conference on Social Robotics*, 548–557. Springer.

141. Hodson, Hal. 2014. *The first family robot*. Elsevier.

142. Miko 3. 2021. Miko 3. *Miko 3*. https://miko.ai/miko.ai. Accessed December 8.

143. BUDDY The Emotional Robot. 2021. *BUDDY The Emotional Robot*. https://buddytherobot.com/en/buddy-the-emotional-robot/. Accessed December 8.

144. Blascovich, Jim. 2002. A theoretical model of social influence for increasing the utility of collaborative virtual environments. In *Proceedings of the 4th international conference on Collaborative virtual environments*, 25–30.

145. Rizzo, Albert, Russell Shilling, Eric Forbell, Stefan Scherer, Jonathan Gratch, and Louis-Philippe Morency. 2016. Autonomous virtual human agents for healthcare information support and clinical interviewing. In *Artificial intelligence in behavioral and mental health care*, 53–79. Elsevier.

146. Bailenson, Jeremy N., Nick Yee, Dan Merget, and Ralph Schroeder. 2006. The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence: Teleoperators and Virtual Environments* 15. MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info …: 359–372.

147. Baylor, Amy L., and Soyoung Kim. 2009. Designing nonverbal communication for pedagogical agents: When less is more. *Computers in Human Behavior* 25. Elsevier: 450–457.

148. Nowak, Kristine L. 2004. The influence of anthropomorphism and agency on social judgment in virtual environments. *Journal of Computer-Mediated Communication* 9. Oxford University Press Oxford, UK: JCMC925.

149. McGuire, William J. 1985. Attitudes and attitude change. *The handbook of social psychology*. Random House: 233–346.

150. Heyselaar, Evelien, Peter Hagoort, and Katrien Segaert. 2017. In dialogue with an avatar, language behavior is identical to dialogue with a human partner. *Behavior research methods* 49. Springer: 46–60.

151. Gong, Li. 2008. How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior* 24. Elsevier: 1494–1509.

152. Kang, Sin-Hwa, and James H. Watt. 2013. The impact of avatar realism and anonymity on effective communication via mobile devices. *Computers in Human Behavior* 29. Elsevier: 1169–1181.

153. Lakoff, George. 2008. *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago press.

154. Lee, Kwan Min, Katharine Liao, and Seoungho Ryu. 2007. Children's responses to computer-synthesized speech in educational media: gender consistency and gender similarity effects. *Human communication research* 33. Oxford University Press Oxford, UK: 310–329.

155. Eastwick, Paul W., and Wendi L. Gardner. 2009. Is it a game? Evidence for social influence in the virtual world. *Social influence* 4. Taylor & Francis: 18–32.

156. Lucas, Gale, Evan Szablowski, Jonathan Gratch, Andrew Feng, Tiffany Huang, Jill Boberg, and Ari Shapiro. 2016. The effect of operating a virtual doppleganger in a 3D simulation. In *Proceedings of the 9th International Conference on Motion in Games*, 167–174.

157. Wauck, Helen, Gale Lucas, Ari Shapiro, Andrew Feng, Jill Boberg, and Jonathan Gratch. 2018. Analyzing the effect of avatar self-similarity on men and women in a search and rescue game. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12.

158. Hall, Judith A., Terrence G. Horgan, and Nora A. Murphy. 2019. Nonverbal communication. *Annual review of psychology* 70. Annual Reviews: 271–294.

159. Brunswik, Egon. 1956. *Perception and the representative design of psychological experiments*. Univ of California Press.

160. Rodriguez-Lujan, Irene, Gonzalo Bailador, Carmen Sanchez-Avila, Ana Herrero, and Guillermo Vidal-de-Miguel. 2013. Analysis of pattern recognition and dimensionality reduction techniques for odor biometrics. *Knowledge-Based Systems* 52. Elsevier: 279–289.

161. Mutic, Smiljana, Eileen M. Moellers, Martin Wiesmann, and Jessica Freiherr. 2016. Chemosensory communication of gender information: Masculinity bias in body odor perception and femininity bias introduced by chemosignals during social perception. *Frontiers in psychology* 6. Frontiers: 1980.

162. Wieser, Matthias J., Paul Pauli, Miriam Grosseibl, Ina Molzow, and Andreas Mühlberger. 2010. Virtual social interactions in social anxiety—the impact of sex, gaze, and interpersonal distance. *Cyberpsychology, Behavior, and Social Networking* 13. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA: 547–554.

163. Zangl, Renate, and Debra L. Mills. 2007. Increased brain activity to infant-directed speech in 6-and 13-month-old infants. *Infancy* 11. Wiley Online Library: 31–62.

164. Ambady, Nalini. 2010. The perils of pondering: Intuition and thin slice judgments. *Psychological Inquiry* 21. Taylor & Francis: 271–278.

165. Zebrowitz, Leslie A., and Mary Ann Collins. 1997. Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and social psychology review* 1. Sage Publications Sage CA: Los Angeles, CA: 204–223.

166. Zaki, Jamil, Jochen Weber, Niall Bolger, and Kevin Ochsner. 2009. The neural bases of empathic accuracy. *Proceedings of the National Academy of Sciences* 106. National Acad Sciences: 11382–11387.

167. Bilakhia, Sanjay, Stavros Petridis, Anton Nijholt, and Maja Pantic. 2015. The MAHNOB Mimicry Database: A database of naturalistic human interactions. *Pattern recognition letters* 66. Elsevier: 52–61.

168. Janssen, Joris H. 2012. A three-component framework for empathic technologies to augment human interaction. *Journal on Multimodal User Interfaces* 6. Springer: 143–161.

169. Van Kleef, Gerben A. 2009. How emotions regulate social life: The emotions as social information (EASI) model. *Current directions in psychological science* 18. SAGE Publications Sage CA: Los Angeles, CA: 184–188.

170. Pentland, Alex. 2007. Social signal processing [exploratory DSP]. *IEEE Signal Processing Magazine* 24. IEEE: 108–111.

171. Poria, Soujanya, Erik Cambria, Amir Hussain, and Guang-Bin Huang. 2015. Towards an intelligent framework for multimodal affective data analysis. *Neural Networks* 63. Elsevier: 104–116.

172. Lanzetta, John T., and Scott P. Orr. 1986. Excitatory strength of expressive faces: Effects of happy and fear expressions and context on the extinction of a conditioned fear response. *Journal of Personality and Social Psychology* 50. American Psychological Association: 190.

173. Vrana, Scott R., Bruce N. Cuthbert, and Peter J. Lang. 1986. Fear imagery and text processing. *Psychophysiology* 23. Wiley Online Library: 247–253.

174. Zheng, Wei-Long, Bo-Nan Dong, and Bao-Liang Lu. 2014. Multimodal emotion recognition using EEG and eye tracking data. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 5040–5043. IEEE.

175. Castellano, Ginevra, Loic Kessous, and George Caridakis. 2008. Emotion recognition through multiple modalities: face, body gesture, speech. In *Affect and emotion in human-computer interaction*, 92–103. Springer.

176. Liu, Zhen-Tao, Fang-Fang Pan, Min Wu, Wei-Hua Cao, Lue-Feng Chen, Jian-Ping Xu, Ri Zhang, and Meng-Tian Zhou. 2016. A multimodal emotional communication based humans-robots interaction system. In *2016 35th Chinese Control Conference (CCC)*, 6363–6368. IEEE.

177. Wu, Chung-Hsien, and Wei-Bin Liang. 2010. Emotion recognition of affective speech based on multiple classifiers using acoustic-prosodic information and semantic labels. *IEEE Transactions on Affective Computing* 2. IEEE: 10–21.

178. Eyben, Florian, Martin Wöllmer, and Björn Schuller. 2009. OpenEAR—introducing the Munich open-source emotion and affect recognition toolkit. In *2009 3rd international conference on affective computing and intelligent interaction and workshops*, 1–6. IEEE.

179. Wilson, Ian. 2008. Using Praat and Moodle for teaching segmental and suprasegmental pronunciation. In *Proceedings of the 3rd international WorldCALL Conference: Using Technologies for Language Learning (WorldCALL 2008)*.

180. Banse, Rainer, and Klaus R. Scherer. 1996. Acoustic profiles in vocal emotion expression. *Journal of personality and social psychology* 70. American Psychological Association: 614.

181. Kleinsmith, Andrea, Nadia Bianchi-Berthouze, and Anthony Steed. 2011. Automatic recognition of non-acted affective postures. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 41. IEEE: 1027–1038.

182. Roether, Claire L., Lars Omlor, Andrea Christensen, and Martin A. Giese. 2009. Critical features for the perception of emotion from gait. *Journal of vision* 9. The Association for Research in Vision and Ophthalmology: 15–15.

183. Burgoon, Judee K., Nadia Magnenat-Thalmann, Maja Pantic, and Alessandro Vinciarelli. 2017. *Social signal processing*. Cambridge University Press.

184. Evers, Vanessa, Nuno Menezes, Luis Merino, Dariu Gavrila, Fernando Nabais, Maja Pantic, Paulo Alvito, and Daphne Karreman. 2014. The development and real-world deployment of FROG, the fun robotic outdoor guide. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, 100–100.

185. Beck, Aryel, Antoine Hiolle, Alexandre Mazel, and Lola Cañamero. 2010. Interpretation of emotional body language displayed by robots. In *Proceedings of the 3rd international workshop on Affective interaction in natural environments*, 37–42.

186. Groff, Ed. 1995. Laban movement analysis: Charting the ineffable domain of human movement. *Journal of Physical Education, Recreation & Dance* 66. Taylor & Francis: 27–30.

187. Lang, Peter J., Mark K. Greenwald, Margaret M. Bradley, and Alfons O. Hamm. 1993. Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology* 30. Wiley Online Library: 261–273.

188. Jatupaiboon, Noppadon, Setha Pan-ngum, and Pasin Israsena. 2013. Emotion classification using minimal EEG channels and frequency bands. In *The 2013 10th international joint conference on Computer Science and Software Engineering (JCSSE)*, 21–24. IEEE.

189. Chen, Xuhai, Jianfeng Yang, Shuzhen Gan, and Yufang Yang. 2012. The contribution of sound intensity in vocal emotion perception: behavioral and electrophysiological evidence. *PLoS one* 7. Public Library of Science San Francisco, USA: e30278.

190. Symons, Ashley E., Wael El-Deredy, Michael Schwartze, and Sonja A. Kotz. 2016. The functional role of neural oscillations in non-verbal emotional communication. *Frontiers in Human Neuroscience* 10. Frontiers: 239.

191. Güntekin, Bahar, and Erol Başar. 2014. A review of brain oscillations in perception of faces and emotional pictures. *Neuropsychologia* 58. Elsevier: 33–51.

192. Başar, Erol, Christina Schmiedt-Fehr, Adile Öniz, and Canan Başar-Eroğlu. 2008. Brain oscillations evoked by the face of a loved person. *Brain research* 1214. Elsevier: 105–115.

193. Jenke, Robert, Angelika Peer, and Martin Buss. 2014. Feature extraction and selection for emotion recognition from EEG. *IEEE Transactions on Affective computing* 5. IEEE: 327–339.

194. Kober, Silvia Erika, Jürgen Kurzmann, and Christa Neuper. 2012. Cortical correlate of spatial presence in 2D and 3D interactive virtual reality: an EEG study. *International Journal of Psychophysiology* 83. Elsevier: 365–374.

195. Pfurtscheller, Gert, A. Stancak Jr, and Ch Neuper. 1996. Event-related synchronization (ERS) in the alpha band—an electrophysiological correlate of cortical idling: a review. *International journal of psychophysiology* 24. Elsevier: 39–46.

196. Levenson, Robert W., and John M. Gottman. 1983. Marital interaction: physiological linkage and affective exchange. *Journal of personality and social psychology* 45. American Psychological Association: 587.

197. Levenson, Robert W., and Anna M. Ruef. 1992. Empathy: a physiological substrate. *Journal of personality and social psychology* 63. American Psychological Association: 234.

198. Wang, Yin, and Susanne Quadflieg. 2015. In our own image? Emotional and neural processing differences when observing human–human vs human–robot interactions. *Social cognitive and affective neuroscience* 10. Oxford University Press: 1515–1524.

199. Rizzolatti, Giacomo, and Laila Craighero. 2004. The mirror-neuron system. *Annu. Rev. Neurosci.* 27. Annual Reviews: 169–192.

200. Frith, Chris D., and Uta Frith. 2006. How we predict what other people are going to do. *Brain research* 1079. Elsevier: 36–46.

201. Urgen, Burcu A., Markus Plank, Hiroshi Ishiguro, Howard Poizner, and Ayse P. Saygin. 2013. EEG theta and Mu oscillations during perception of human and robot actions. *Frontiers in neurorobotics* 7. Frontiers: 19.

202. Yoon, Sue, Maryam Alimardani, and Kazuo Hiraki. 2021. The Effect of Robot-Guided Meditation on Intra-Brain EEG Phase Synchronization. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 318–322.

203. Cafaro, Angelo, Johannes Wagner, Tobias Baur, Soumia Dermouche, Mercedes Torres Torres, Catherine Pelachaud, Elisabeth André, and Michel Valstar. 2017. The NoXi database: multimodal recordings of mediated novice-expert interactions. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, 350–359.

204. McKeown, Gary, Michel Valstar, Roddy Cowie, Maja Pantic, and Marc Schroder. 2011. The semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE transactions on affective computing* 3. IEEE: 5–17.

205. Douglas-Cowie, Ellen, Roddy Cowie, Ian Sneddon, Cate Cox, Orla Lowry, Margaret Mcrorie, Jean-Claude Martin, Laurence Devillers, Sarkis Abrilian, and Anton Batliner. 2007. The HUMAINE database: Addressing the collection and annotation of naturalistic and induced emotional data. In *International conference on affective computing and intelligent interaction*, 488–500. Springer.

206. Koelstra, Sander, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. 2011. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing* 3. IEEE: 18–31.

207. Ringeval, Fabien, Andreas Sonderegger, Juergen Sauer, and Denis Lalanne. 2013. Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. In *2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG)*, 1–8. IEEE.

208. Lefter, Iulia, Catholijn M. Jonker, Stephanie Klein Tuente, Wim Veling, and Stefan Bogaerts. 2017. NAA: A multimodal database of negative affect and aggression. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, 21–27. IEEE.

209. Newman, Benjamin A., Reuben M. Aronson, Siddhartha S. Srinivasa, Kris Kitani, and Henny Admoni. 2022. HARMONIC: A multimodal dataset of assistive human–robot collaboration. *The International Journal of Robotics Research* 41. SAGE Publications Sage UK: London, England: 3–11.

210. Hazer-Rau, Dilana, Sascha Meudt, Andreas Daucher, Jennifer Spohrs, Holger Hoffmann, Friedhelm Schwenker, and Harald C. Traue. 2020. The uulmMAC database—A multimodal affective corpus for affective computing in human-computer interaction. *Sensors* 20. Multidisciplinary Digital Publishing Institute: 2308.

211. Rasheed, Umer, Yasir Tahir, Shoko Dauwels, Justin Dauwels, Daniel Thalmann, and Nadia Magnenat-Thalmann. 2013. Real-time comprehensive sociometrics for two-person dialogs. In *International Workshop on Human Behavior Understanding*, 196–208. Springer.

212. Moussa, Maher Ben, Zerrin Kasap, Nadia Magnenat-Thalmann, Krishna Chandramouli, Seyed Navid Haji Mirza, Qianni Zhang, Ebroul Izquierdo, Iordanis Biperis, and Petros Daras. 2010. Towards an expressive virtual tutor: an implementation of a virtual tutor based on an empirical study of non-verbal behaviour. In *Proceedings of the 2010 ACM workshop on Surreal media and virtual cloning*, 39–44.

213. Giger, Jean-Christophe, Nuno Piçarra, Patrícia Alves-Oliveira, Raquel Oliveira, and Patrícia Arriaga. 2019. Humanization of robots: Is it really such a good idea? *Human Behavior and Emerging Technologies* 1. Wiley Online Library: 111–123.

214. Libin, Alexander V., and Elena V. Libin. 2004. Person-robot interactions from the robopsychologists' point of view: The robotic psychology and robotherapy approach. *Proceedings of the IEEE* 92. IEEE: 1789–1803.

215. Hong, Alexander, Nolan Lunscher, Tianhao Hu, Yuma Tsuboi, Xinyi Zhang, Silas Franco dos Reis Alves, Goldie Nejat, and Beno Benhabib. 2020. A Multimodal Emotional Human-Robot Interaction Architecture for Social Robots Engaged in Bidirectional Communication\vspace* 7pt. *IEEE transactions on cybernetics*. IEEE.

216. Ficocelli, Maurizio, Junichi Terao, and Goldie Nejat. 2015. Promoting interactions between humans and robots using robotic emotional behavior. *IEEE transactions on cybernetics* 46. IEEE: 2911–2923.

217. Paletta, Lucas, Maria Fellner, Sandra Schüssler, Julia Zuschnegg, Josef Steiner, Alexander Lerch, Lara Lammer, and Dimitrios Prodromou. 2018. AMIGO: Towards social robot based motivation for playful multimodal intervention in dementia. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*, 421–427.

218. Nomura, Tatsuya, Takayuki Kanda, Tomohiro Suzuki, and Kensuke Kato. 2008. Prediction of human behavior in human--robot interaction using psychological scales for anxiety and negative attitudes toward robots. *IEEE transactions on robotics* 24. IEEE: 442–451.

219. Ratajczyk, Dawid, Marcin Jukiewicz, and Pawel Lupkowski. 2019. Evaluation of the uncanny valley hypothesis based on declared emotional response and psychophysiological reaction. *Bio-Algorithms and Med-Systems* 15. De Gruyter.

220. Moreau, Quentin, Matteo Candidi, Vanessa Era, Gaetano Tieri, and Salvatore Maria Aglioti. 2019. Frontal and occipito-temporal Theta activity as marker of error monitoring in Human-Avatar joint performance. *BioRxiv*. Cold Spring Harbor Laboratory: 402149.

221. Amershi, Saleema, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, and Kori Inkpen. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*, 1–13.

222. Minderer, Matthias, Christopher D. Harvey, Flavio Donato, and Edvard I. Moser. 2016. Virtual reality explored. *Nature* 533. Nature Publishing Group: 324–325.

223. Stavroulia, Kalliopi Evangelia, Evangelia Baka, Andreas Lanitis, and Nadia Magnenat-Thalmann. 2018. Designing a virtual environment for teacher training: Enhancing presence and empathy. In *Proceedings of Computer Graphics International 2018*, 273–282.

224. Von Stein, Astrid, and Johannes Sarnthein. 2000. Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta synchronization. *International journal of psychophysiology* 38. Elsevier: 301–313.

225. Abhang, Priyanka A., Bharti W. Gawali, and Suresh C. Mehrotra. 2016. Technological basics of EEG recording and operation of apparatus. *Introduction to EEG-and Speech-Based Emotion Recognition*. Academic Press: 19–50.

226. da Silva, Fernando Lopes. 1991. Neural mechanisms underlying brain waves: from neural membranes to networks. *Electroencephalography and clinical neurophysiology* 79. Elsevier: 81–93.

227. Regenbrecht, Holger, and Thomas Schubert. 2002. Real and illusory interactions enhance presence in virtual environments. *Presence: Teleoperators & Virtual Environments* 11. MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info …: 425–434.

228. Argento, Emanuele, George Papagiannakis, Eva Baka, Michail Maniadakis, Panos Trahanias, Michael Sfakianakis, and Ioannis Nestoros. 2017. Augmented Cognition via Brainwave Entrainment in Virtual Reality: An Open, Integrated Brain Augmentation in a Neuroscience System Approach. *Augmented Human Research* 2. Springer: 3.

229. Babiloni, Claudio, Claudio Del Percio, Fabrizio Vecchio, Fabio Sebastiano, Giancarlo Di Gennaro, Pier P. Quarato, Roberta Morace, Luigi Pavone, Andrea Soricelli, and Giuseppe Noce. 2016. Alpha, beta and gamma electrocorticographic rhythms in somatosensory, motor, premotor and prefrontal cortical areas differ in movement execution and observation in humans. *Clinical Neurophysiology* 127. Elsevier: 641–654.

230. Cavanagh, James F., and Michael J. Frank. 2014. Frontal theta as a mechanism for cognitive control. *Trends in cognitive sciences* 18. Elsevier: 414–421.

231. Baka, Evangelia, Kalliopi Evangelia Stavroulia, Nadia Magnenat-Thalmann, and Andreas Lanitis. 2018. An EEG-based evaluation for comparing the sense of presence between virtual and physical environments. In *Proceedings of Computer Graphics International 2018*, 107–116.

232. Pfefferbaum, Betty, and Carol S. North. 2020. Mental health and the Covid-19 pandemic. *New England Journal of Medicine* 383. Mass Medical Soc: 510–512.

233. Delorme, Arnaud, and Scott Makeig. 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods* 134. Elsevier: 9–21.

234. Frantzidis, Christos A., Aristea-Kiriaki I. Ladas, Ana B. Vivas, Magda Tsolaki, and Panagiotis D. Bamidis. 2014. Cognitive and physical training for the elderly: evaluating outcome efficacy by means of neurophysiological synchronization. *International Journal of Psychophysiology* 93. Elsevier: 1–11.

235. Frantzidis, Christos A., Charalampos Bratsas, Christos L. Papadelis, Evdokimos Konstantinidis, Costas Pappas, and Panagiotis D. Bamidis. 2010. Toward emotion aware computing: an integrated approach using multichannel neurophysiological recordings and affective visual stimuli. *IEEE transactions on Information Technology in Biomedicine* 14. IEEE: 589–597.

236. Quiroga, R. Quian, and M. Schürmann. 1999. Functions and sources of event-related EEG alpha oscillations studied with the Wavelet Transform. *Clinical Neurophysiology* 110. Elsevier: 643–654.

237. Sapiński, Tomasz, Dorota Kamińska, Adam Pelikant, and Gholamreza Anbarjafari. 2019. Emotion recognition from skeletal movements. *Entropy* 21. Multidisciplinary Digital Publishing Institute: 646.

238. Mishra, Nidhi, Evangelia Baka, and Nadia Magnenat Thalmann. 2021. Exploring Potential and Acceptance of Socially Intelligent Robot. In *Intelligent Scene Modeling and Human-Computer Interaction*, 259–282. Springer.

239. EMG and Motion Tools. 2015. *Cometa Systems*. July 6.

240. Reaz, Mamun Bin Ibne, M. Sazzad Hussain, and Faisal Mohd-Yasin. 2006. Techniques of EMG signal analysis: detection, processing, classification and applications. *Biological procedures online* 8. Springer: 11–35.

241. Gupta, Ashutosh, Tabassum Sayed, Ridhi Garg, and Richa Shreyam. 2017. EMG signal analysis of healthy and neuropathic individuals. In *IOP Conference Series: Materials Science and Engineering*, 225:012128. IOP Publishing.

242. Fukuda, Thiago Yukio, Jorge Oliveira Echeimberg, José Eduardo Pompeu, Paulo Roberto Garcia Lucareli, Silvio Garbelotti, Rafaela Okano Gimenes, and Adilson Apolinário. 2010. Root mean

square value of the electromyographic signal in the isometric torque of the quadriceps, hamstrings and brachial biceps muscles in female subjects. *J Appl Res* 10: 32–39.

243. Ferrari, E., G. Cooper, N. D. Reeves, and E. F. Hodson-Tole. 2018. Surface electromyography can quantify temporal and spatial patterns of activation of intrinsic human foot muscles. *Journal of Electromyography and Kinesiology* 39. Elsevier: 149–155.

244. Crumpton, Joe, and Cindy L. Bethel. 2016. A survey of using vocal prosody to convey emotion in robot speech. *International Journal of Social Robotics* 8. Springer: 271–285.

245. Trigeorgis, George, Fabien Ringeval, Raymond Brueckner, Erik Marchi, Mihalis A. Nicolaou, Björn Schuller, and Stefanos Zafeiriou. 2016. Adieu features? end-to-end speech emotion recognition using a deep convolutional recurrent network. In *2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 5200–5204. IEEE.

246. Livingstone, Steven R., and Frank A. Russo. 2018. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PloS one* 13. Public Library of Science San Francisco, CA USA: e0196391.

247. McFee, Brian, Colin Raffel, Dawen Liang, Daniel PW Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. 2015. librosa: Audio and music signal analysis in python. In *Proceedings of the 14th python in science conference*, 8:18–25. Citeseer.

248. Giannakopoulos, Theodoros. 2015. pyaudioanalysis: An open-source python library for audio signal analysis. *PloS one* 10. Public Library of Science San Francisco, CA USA: e0144610.

249. Goutte, Cyril, and Eric Gaussier. 2005. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *European conference on information retrieval*, 345–359. Springer.

250. Watson, David, and Lee Anna Clark. 1999. The PANAS-X: Manual for the positive and negative affect schedule-expanded form.

251. Abhang, Priyanka A., Bharti W. Gawali, and Suresh C. Mehrotra. 2016. Technical aspects of brain rhythms and speech parameters. *Introduction to EEG-and Speech-Based Emotion Recognition*. Elsevier: 51–79.

252. Teixeira, João Paulo, Carla Oliveira, and Carla Lopes. 2013. Vocal acoustic analysis–jitter, shimmer and hnr parameters. *Procedia Technology* 9. Elsevier: 1112–1122.

253. de Felippe, Ana Clara Naufel, Maria Helena Marotti Martelletti Grillo, and Thaís Helena Grechi. 2006. Standardization of acoustic measures for normal voice patterns. *Brazilian journal of otorhinolaryngology* 72. Elsevier: 659–664.

254. Yumoto, Eiji, Wilbur J. Gould, and Thomas Baer. 1982. Harmonics-to-noise ratio as an index of the degree of hoarseness. *The journal of the Acoustical Society of America* 71. Acoustical Society of America: 1544–1550.

255. Syrdal, Dag Sverre, Kerstin Dautenhahn, Kheng Lee Koay, and Michael L. Walters. 2009. The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study. *Adaptive and emergent behaviour and complex systems*. SSAISB.

256. Alekseichuk, Ivan, Zsolt Turi, Gabriel Amador de Lara, Andrea Antal, and Walter Paulus. 2016. Spatial working memory in humans depends on theta and high gamma synchronization in the prefrontal cortex. *Current Biology* 26. Elsevier: 1513–1521.

257. Prathaban, Sachin, Vishal Sisodia, and Sadasivan Puthusserypady. 2019. Consequence of Stress on Cognitive Performance: An EEG and HRV Study. In *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*, 1969–1974. IEEE.

258. Louis, Erik K. St, Lauren C. Frey, Jeffrey W. Britton, Jennifer L. Hopp, Pearce Korb, Mohamad Z. Koubeissi, William E. Lievens, and Elia M. Pestana-Knight. 2016. The normal eeg. *Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants [Internet]*. American Epilepsy Society.

259. Biau, Emmanuel, and Sonja A. Kotz. 2018. Lower beta: A central coordinator of temporal prediction in multimodal speech. *Frontiers in human neuroscience* 12. Frontiers: 434.

260. Christofi, Maria, Evangelia Baka, Kalliopi-Evangelia Stavroulia, Despina Michael-Grigoriou, Andreas Lanitis, and Nadia Magnenat-Thalmann. 2018. Studying Levels of Presence in a Virtual Environment Simulating Drug Use in Schools: Effect on Different Character Perspectives. In *ICAT-EGVE*, 163–170.

261. Agisoft Metashape. 2021. https://www.agisoft.com/. Accessed December 17.

262. Visage Technologies - Face tracking, analysis and recognition technology. 2021. *Visage Technologies*. https://visagetechnologies.com/. Accessed December 17.