

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

School of Computer Science and Statistics

Evaluating the Influence of Vocal and Visual Realism on Subjective Experiences of Empathy for Virtual Characters

Darragh Higgins

Supervisor: Associate Prof. Rachel McDonnell

January 24, 2025

A thesis submitted in partial fulfilment
of the requirements for the degree of
Doctorate in Computer Science

Abstract

With the expanding range of use cases for advancements in animation and voice synthesis, more opportunities arise for interactions with realistic virtual characters. As yet, it is unclear to what extent realism, either in appearance or voice, influences our reactions to emotional expressions in such characters.

This work describes the roles of vocal and visual realism in providing opportunities for empathy. These dimensions of realism are used to convey emotional expressions in faces and voices of virtual characters. In this work, physiological behaviour is observably influenced by these expressions in a manner that is consistent with the phenomenon of empathy. Moreover these dimensions of realism are seen to impact self-reports of empathy for these characters.

Over the last four years, our experiments have steadily narrowed in on aspects of virtual characters which influence perception, cognition and emotion in users. While our studies originally focused on the effect of appearance and the role of individual user traits, it gradually became clear that bimodal realism (i.e. vocal and visual realism) was the topic which required attention in this domain.

With this more holistic approach to realism in virtual characters, the question arose as to whether these variations could affect empathic responses. Since empathy, as measured in this thesis, was stimulated by the perception of emotions that varied in intensity between levels of realism, it became necessary to question whether visual or vocal realism may provide 'affordances' for emotion perception and empathy, as described in the language of ecological psychology. This question directed our research towards evaluations of of vocal-visual realism combinations that were ultimately observed to influence both physiological measurements of motor empathy, and self-reported perceptions of cognitive, affective and compassionate empathy.

What follows in this thesis is an attempt to embed these findings into recent trends in computer graphics and cognitive science research. Each research project that was conducted over the course of this doctorate is explained with reference to these fields, and their unifying domain of human-computer interaction, as well as the other fields that they invariably touch off. The findings ultimately indicate that realism and emotional expressions in virtual characters present affordance opportunities that blur the distinction between object-perception and social behaviour.

Acknowledgements

In truth, the list of names in this acknowledgments section should be at least as long as the thesis itself. Through all the difficult times of the last four years, countless wonderful people have benefited this work to a significant degree, in ways that I cannot truly express.

I must give special mention to my primary supervisor, Dr. Rachel McDonnell, and my external supervisor, Dr. Benjamin Cowan, who not only gave me every opportunity to complete this work but in the process broadened my mind, knowledge and skill-set in ways that will benefit me forever. I am profoundly grateful for the chances they have given me and the resources they have both provided for me to conduct this research.

Two postdoctoral researchers who have featured heavily in this work are Dr. Rebecca Fribourg and Dr. Dalila Burin. Together with my supervisors, their incredible skill, dilligence, patience, wisdom and insight have been invaluable to me throughout this entire process.

I also must thank my examiners, Dr. Carol O'Sullivan and Dr. Catherine Pelachaud. I have composed this work based on the impressive works that they have been a part of during their careers and hope that they enjoy reading it the way that I have greatly enjoyed and benefited from reading theirs.

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Some of the work presented in this thesis includes both unpublished and published work of others.

_____

Darragh Higgins

Contents

1 Introduction **1**

1.1 Motivation . . . 5
1.2 Scope and Limitations . . . 6
1.3 Thesis Overview . . . 7
1.4 Contributions . . . 9

2 Background **10**

2.1 Animated Virtual Characters . . . 11
2.1.1 Avatars and Embodiment . . . 13
2.1.2 Virtual Humans . . . 18
2.1.3 Subjective Experiences of Virtual Characters . . . 21
2.2 Emotional Affordances . . . 27
2.2.1 Models of Emotion . . . 27
2.2.2 Affordances . . . 32
2.2.3 Empathy and Direct Perception . . . 35
2.2.4 Face and Voice Processing . . . 37
2.3 Perception of Realism . . . 42
2.3.1 Visual Realism . . . 46
2.3.2 Vocal Realism . . . 49
2.3.3 The Uncanny Valley . . . 51
2.4 Methodologies to Assess Empathy for Virtual Characters . . . 57
2.4.1 Valence and Virtual Characters . . . 57
2.4.2 Empathy in Research . . . 59
2.4.3 Physiological Correlates of Empathy . . . 60
2.4.4 Summary . . . 63

3 Experiments 1A-C: Trait Effects on Experiences of Humanoid Avatars **66**

3.1 Experiment 1A: Pilot Study of Emotion Communication . . . 67

3.1.1 Stimuli Creation . . . 67
3.1.2 Experiment Design . . . 70
3.1.3 Results . . . 73
3.1.4 Discussion for Experiment 1A . . . 77
3.2 Experiment 1B: Main Experiment of Emotion Communication . . . 78
3.2.1 Stimuli Creation . . . 79
3.2.2 Experiment Design . . . 80
3.2.3 Measurements . . . 82
3.2.4 Hypotheses . . . 83
3.2.5 Results . . . 84
3.2.6 Discussion for Experiment 1B . . . 90
3.2.7 General Discussion for Experiments 1A and 1B . . . 92
3.3 Experiment 1C: Subjective Perceptions of Human and Non-human Avatars 95
3.4 Experiment . . . 96
3.4.1 Participants . . . 96
3.4.2 Protocol . . . 96
3.4.3 Measurements . . . 98
3.5 Analysis . . . 99
3.6 Results . . . 100
3.6.1 Analysis of Variance . . . 100
3.6.2 Correlations . . . 103
3.7 Discussion for Experiment 1C . . . 105
3.8 Conclusions . . . 110

4 Experiments 2A-C: The Effect of Unimodal Realism Variations on the Perception of Virtual Humans 112
4.1 Experiment 2A: Visual Realism Variations with Real Voices . . . 112
4.1.1 Stimuli and Procedure . . . 113
4.1.2 Results . . . 115
4.1.3 Discussion: Experiment 2A . . . 116
4.2 Experiment 2B: Synthetic Voices and Visual Consistency . . . 117
4.2.1 Stimuli . . . 119
4.2.2 Measures . . . 120
4.2.3 Participants and Procedure . . . 121
4.2.4 Analysis . . . 122
4.2.5 Results . . . 122
4.2.6 Discussion: Experiment 2B . . . 125
4.3 Experiment 2C: Animated Facial Expressions and Empathic Responses . . . 126

4.3.1 Stimuli and Procedure . . . 127
4.3.2 Results . . . 129
4.3.3 Discussion: Experiment 2C . . . 132
4.4 Conclusions, intuitions and planned developments . . . 133

5 Experiment 3: Examining Empathic and Physiological Responses to Bimodal Realism in Emotionally Expressive Virtual Humans 135
5.1 Hypotheses . . . 136
5.2 Materials and Methods . . . 138
5.2.1 Stimuli Synthesis . . . 138
5.2.2 Participants . . . 141
5.2.3 Measures . . . 141
5.2.4 Procedure . . . 144
5.3 Results . . . 146
5.3.1 EMG Hypotheses (H1 a-c) . . . 147
5.3.2 ET Hypotheses (H2 a-c) . . . 148
5.3.3 State Empathy and Uncanny Valley Hypotheses (H3, H4a-b) . . . 151
5.3.4 Exploratory Analysis . . . 153
5.3.5 Cross-Experimental Analysis . . . 155
5.4 Discussion . . . 158
5.5 Conclusion . . . 163

6 Conclusions 165
6.1 Contributions . . . 165
6.2 Limitations and Future Work . . . 169
6.3 Conclusions . . . 174

List of Figures

1.1 Summary of main research focus: How do users respond empathically to emotions afforded by virtual characters at different levels of realism? . . . 2

2.1 Example of Avatar realism conditions from Fribourg et al and their experiment on point of view [Fribourg et al., 2020] . . . 14
2.2 The sensorimotor contingencies model proposed by [Gonzalez-Franco and Lanier, 2017] which is alleged to describe the mechanisms by which virtual reality illusions take place. . . . 15
2.3 Personalised avatars used in the experimental work of Waltemate et al [Waltemate et al., 2018] . . . 16
2.4 Face only avatars used to by Oh et al to test the effects of enhanced smiling behaviour [Bailenson et al., 2016] . . . 18
2.5 Components of a virtual human body [Burden and Savin-Baden, 2019] . 19
2.6 Virtual humans presented at different degrees of visual realism, as employed by Parmar et al [Parmar et al., 2022] . . . 21
2.7 Avatars at different levels of attractiveness have been shown to influence user behaviour [Yee and Bailenson, 2007] . . . 23
2.8 Dewez et al 2019 assessed the relationship between Five Factor personality traits and virtual embodiment . . . 25
2.9 The facial action coding system applies numeric values to muscle movements in the human face that are used for facial expressions of emotion [Hassouneh et al., 2020] . . . 29
2.10 The continuum of affordances as described by Morie et al in the study of affordances in virtual environments [Morie et al., 2005] . . . 33
2.11 A useful schematic for describing the different ways an emotion may be expressed and perceived [Lange et al., 2022] . . . 38
2.12 Dummy and real faces used to study the neural responses to the social affordances provided by faces [Debruille et al., 2012] . . . 39

2.13 Types of realism in the animation of virtual characters as defined by Rekik et al [Rekik et al., 2024] 46
2.14 Example of differing levels of visual realism, used by Schindler et al [Schindler et al., 2017] to study realism and emotion perception. 48
2.15 Virtual characters in human and robot form tested in conjunction with real voices and text-to-speech generated voices [Ferstl et al., 2021a] . . . 51
2.16 The proposed dimensions of the Uncanny Valley 52
2.17 Some of the animated characters used by Ho & MacDorman to develop metrics of the Uncanny Valley [Ho and MacDorman, 2010a] 59
2.18 Corrugator supercilii and Zygomaticus major face muscles that are commonly associated with negative and positive emotional valence respectively [Martínez-Fiestas et al., 2015] 62

3.1 Pilot Experiment: Some examples of the customised avatars created by the participants. 68
3.2 Video call between participant (left) and experimenter (right), both embodied in virtual avatars and viewed side-by-side on-screen. 70
3.3 Pilot Experiment: Results for questions Emotional Valence, Enfacement, Proteus Effect and Social Presence. X-Axis labels are the variable names from Table 3.1. Star indicates statistical significance < 0.05 against the reference value '4'. 74
3.4 Main Experiment: Some examples of the customised avatars created by the participants using the photo avatar creator tool. 80
3.5 Main Experiment: Boxplots of all the dependent variables depending on factor Emotion for *Avatar* condition (A) and *Voice* condition (B). 86
3.6 Main Experiment: Summary of significant effects of Emotion and PartnerRepresentation on Agency1 (A), SharedPresence1 (B), SharedPresence2 (C), SocialPresence1 (D) and Valence (E). 90
3.7 Office scene for experiment with human avatar. The side mirror was included to show participants the movements of the avatar as mapped to their own movements. 97
3.8 Three avatars embodied during the experiment. Participants chose either the female (right) or male (left) depending on their own gender, and were also embodied the the non-human avatar (center). The avatars were scaled to the size of each individual participant. 98
3.9 Example of shape trace task from the user's point of view. 99
3.10 Estimated marginal means for Avatar, Priming and Ownership. Human avatar (blue) evoked significantly more ownership than non-human (red). 100

3.11 Plotted correlation coefficients for primed group personality traits with SoE and SMS results. Significant results labelled with a * for both human (h) and non-human (nh) avatar conditions. 101

4.1 Characters 1–6. Range of male (left) and female (right) virtual humans used in experiment, shown in the high fidelity condition. 113
4.2 Screenshots from scenarios (left) Sad, (middle) Friendly, and (right) Unfriendly for character in Experiment 2. Participants viewed the character's performance with either a natural, synthetic or text-to-speech voice. . . . 120
4.3 Our experiment investigated empathy towards a range of emotional virtual characters. Shown are two of the six virtual characters used in the study, at high visual realism (left) and low visual realism (right). Screenshots from the positive emotion animation are shown on top and negative on the bottom. 128
4.4 Estimated marginal means for Cognitive Empathy in all three emotion conditions 130
4.5 Estimated marginal means for Affective Empathy in all three emotion conditions 131
4.6 Estimated marginal means for Compassionate Empathy in all three emotion conditions 131

5.1 Two of the character models used in the study, expressing happiness (top) and sadness (bottom) at both high (left) and low (right) levels of visual realism. 139
5.2 Electrode placement for measurement of Zygomaticus major (bottom) and Corrugator Supercilii (top) activity 142
5.3 Pilot run of participant setup for eye-tracking measurement. The Tobii bar was placed below the stimuli which were presented on screen. 143
5.4 Schematic of experiment procedure. 145
5.5 Bar plot of mean RMS for Zygomaticus activity across all conditions, with standard error bars. 149
5.6 Bar plot of mean RMS for Corrugator activity across Happy (left) and Sad (right) conditions, including standard error bars. 150
5.7 ET areas of interest (AOIs) shown on high appearance realism character (white male), eyes (highlighted in pink), nose (highlighted in yellow), and mouth (highlighted in orange) 151

5.8 Box plots for eye fixation count across Happy and Sad emotion conditions, for all Areas of Interest. Significantly higher fixation count was observed on all Areas of Interest for Happy (left) compared to Sad (Right). 152
5.9 Bar plots for reported state empathy across voice conditions. 153
5.10 Box plots for reported cognitive empathy across voice and appearance conditions. 156
5.11 Box plots for reported affective empathy across voice and appearance conditions. 157
5.12 Box plots for reported compassionate empathy across voice and appearance conditions. 157

List of Tables

3.1 Pilot Experiment: Questions arranged by group and variable name. Each question could be answered on a scale from 1 – "Not at all" to 7 – "Extremely". 71

3.2 Main Experiment: Questions for the main experiment by variable name. Each question could be answered on a scale from 1 – "Strongly Disagree" to 5 – "Strongly Agree". Question with star were only asked in the *Avatar* condition. 83

3.3 Main Experiment: Summary of main effects on subjective scores 85

3.4 Main Experiment: Spearman correlations between personality scores and questionnaire items. 89

4.1 Questions arranged by variable name. Each question could be answered on a scale from 1 – "Strongly Disagree" to 7– "Strongly Agree". Question with star were only asked in the *Avatar* condition. 121

4.2 Summary of all main effects with corresponding post–hoc analysis for Experiment 2. 123

5.1 Questionnaire assessed at the end of each condition (block). The first columns represent the measures underlying each statement of the questionnaire (not disclosed to participants); the third column describes the actual statement that participants rated on a 1-7 Likert scale (not ART transformed used in the analysis). The last eight columns represent the averaged (SD) values between participants for each condition. 147

5.2 Descriptive statistics of the RMS recorded with the EMG. The first line refers to the conditions, the second and third are the RMS values for corrugator supercilii (Cor) and zygomaticus major (Zyg) respectively. Values are reported as average (SD). 148

5.3 Descriptive statistics of the fixation durations recorded with the ET, expressed in seconds. The first line refers to the conditions, the second, third and fourth are the values for the three AOIs, eyes, mouth and nose, respectively. Values are reported as average (SD). 155
5.4 Descriptive statistics of the fixation count recorded with the ET, expressed in n. The first line refers to the conditions, the second, third and fourth are the values for the three AOIs, eyes, mouth and nose, respectively. Values are reported as average (SD). 155

Relevant Publications

Journal

Higgins, D., Fribourg, R., & McDonnell, R. (2021). Remotely perceived: Investigating the influence of valence on self-perception and social experience for dyadic video-conferencing with personalized avatars. Frontiers in Virtual Reality, 2, 668499.

Higgins, D., Zibrek, K., Cabral, J., Egan, D., & McDonnell, R. (2022). Sympathy for the digital: Influence of synthetic voice on affinity, social presence and empathy for photorealistic virtual humans. Computers & Graphics, 104, 116-128.

Conference

Higgins, D., Egan, D., Fribourg, R., Cowan, B., & McDonnell, R. (2021, September). Ascending from the valley: Can state-of-the-art photorealism avoid the uncanny?. In ACM Symposium on Applied Perception 2021 (pp. 1-5).

Higgins, D., Zhan, Y., Cowan, B. R., & McDonnell, R. (2023, August). Investigating the effect of visual realism on empathic responses to emotionally expressive virtual humans. In ACM Symposium on Applied Perception 2023 (pp. 1-7).

Higgins, D., Fribourg R., Normand, J., & McDonnell, R. (2024, December). Priming and personality effects on the sense of embodiment for human and non-human avatars in virtual reality. In International Conference on Artificial Reality and Telexistence & the 29th Eurographics Symposium on Virtual Environments (pp 1-11).

In Review

Higgins, D., Cowan, B., & McDonnell, R. (2024). Evaluating empathic responses to bimodal realism in emotionally expressive virtual humans: an eye-tracking and facial electromyography study.

Posters

Higgins, D., & McDonnell, R. (2021, March). A preliminary investigation of avatar use in video-conferencing. In 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) (pp. 540-541). IEEE.

1 Introduction

This thesis describes doctoral work dedicated to furthering our understanding of the effects of virtual character realism on human attitudes and behaviours. Specifically, the main research question addressed in this thesis is: How are our emotional interactions with virtual representations of humans influenced by vocal and visual realism? A secondary topic addressed by this research is: could the influences of vocal and visual realism on emotion perception be subject to the influence of individual trait differences in users? These questions are developed from the basis of several experiments, each geared towards answering related research questions of their own. Each of these sits at the intersection of a number of different disciplines, which are addressed in detail throughout this work.

In order to address and interpret answers to these questions, it is important to reduce them to measurable hypotheses, sub-questions, and graspable concepts. Hence, it is necessary to give a thorough definition of the terms used in this thesis: Evaluating the Influence of Vocal and Visual Realism on Subjective Experiences of Empathy for Virtual Characters.

Virtual characters, in this work, are considered as computer generated character models that resemble real humans, that may be used to represent both human users and non-playable characters in digital environments.

The concept of empathy that is employed in this thesis encompasses several different definitions (although not an exhaustive list) of empathy. Empathy can be used to mean a variety of different things, including sympathy, or understanding, or feeling what someone else is feeling.

We chose to cast a wide net over the use of the term empathy, which included the ability to recognise positive and negative emotions in virtual characters, understand them, and respond to them. Within these definitions, which are outlined in depth during Chapter 2, this thesis addressed:

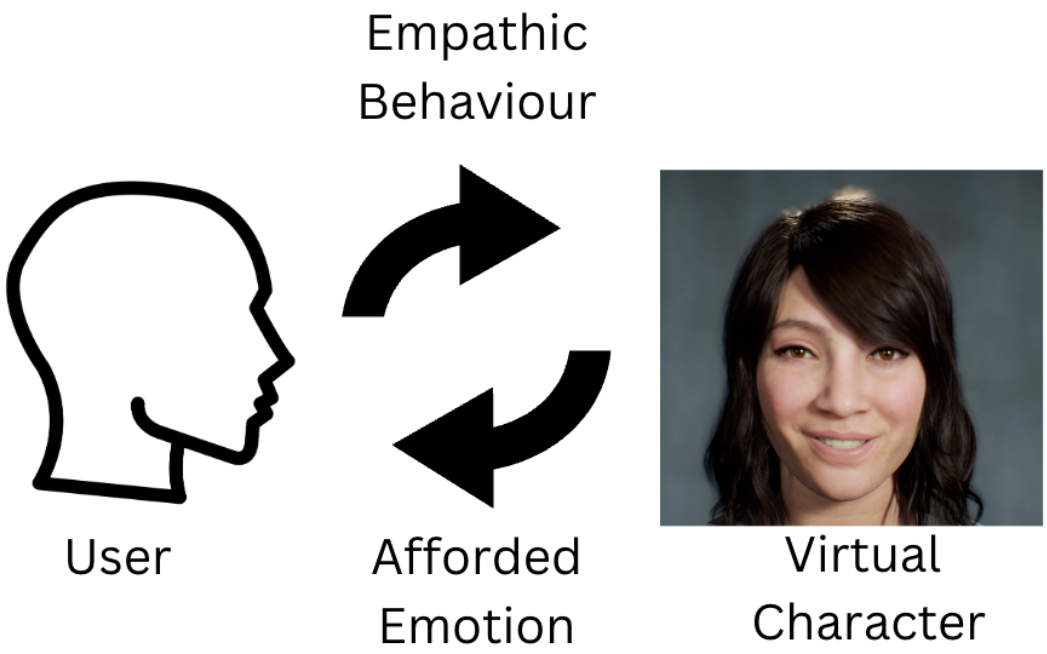


Figure 1.1: Summary of main research focus: How do users respond empathically to emotions afforded by virtual characters at different levels of realism?

- Emotion recognition - how do we interpret positive and negative emotions as they are represented by a character perspective?
- Emotion response - does the recognition of emotion in virtual characters change how we feel, understand or move?

One core assumption of this work is that different people respond to the same stimulation of empathy in different ways, depending on their subjective experiences.

This assumption required the use of validated tools to capture different aspects of our experimental participants, including measurable dimensions of personality and individual capacity for empathy, while ensuring validity by including dimensions of gender and ethnicity. This 'subjective' viewpoint was an important consideration in the variation of the characters we employed, in addition to their dimensions of realism.

The two kinds of realism that we focused on were visual and vocal realism. To be clear, the definitions we eventually focused on were:

- Visual realism: how much does the virtual character look like a real human?
- Vocal realism: and how much does the character's voice sound like a real human

voice?

Within this scope, another assumption of this thesis is that different levels of visual and vocal realism *afford* different possibilities for emotion recognition and response. For example, depictions of key face cues of emotion, such as movement in the brow or mouth, can be considered as affordances that allow for emotion perception in observers. Based on the literature, it was apparent that the full extent of the possible influences of realism in the perception of emotion from such afforded cues remained to be definitively understood.

While the methods of evaluation and documented influences will be explained in depth throughout this thesis, the studies on which they are based ultimately show that the perception of emotion in realistic virtual characters involves empathy, in terms of both physiological and self-reported responses. In the main experiment of this thesis, the findings show that there are some differences between realistic affordances for empathy and unrealistic affordances for empathy; namely that realistic human voices invite higher levels of cognitive empathy compared to synthetic voices, which are less useful for translating emotional information in virtual characters at differing degrees of visual realism. In addition, there is evidence that vocal and visual affordances combined can be influential at the physiological level, to a greater degree under positive emotion conditions.

The experiments described here were designed to capture a wide range of effects that emerge when subjects are presented with differing modes of realism and positive or negative emotional contexts, to measure and assess the overall influence they have on perception of virtual characters. The studies described in Chapter 3 were conducted with a view to evaluating perceived embodiment, human-likeness and facial realism in avatar interactions, under varied conditions of emotional valence. This is to be distinguished from the virtual human perception experiments of Chapters 4 and 5, where independent virtual human stimuli were created as a means to measure the influence of voice and appearance realism as observably influential in this separate class of digital interactions.

The methods and designs which informed these experiments have been adapted from the relevant literature, and employed in a novel way to provide new insights into the experiences of empathy for virtual characters. This research can be considered to benefit a wide range of fields that wish to study emotion, realism and virtual characters.

It is important to state from the outset that there were numerous challenges present in the conduct of this work. As such, there are significant limitations to the experimental reports that are presented, which will be outlined in the coming sections. However, as shall be made clear, the data gathered throughout four years of doctoral research have

nonetheless provided valuable contributions to the fields of human-computer interaction, computer graphics and cognitive science.

1.1 Motivation

Virtual characters bear increasing importance in the modern technological society. Indeed, prior to the commencement of the following works, there were already significant advancements made in our understanding of virtual characters. These advancements have been utilised in film and television, as well as games, training simulations and physical rehabilitation applications.

While the technical development of these characters improves drastically with each passing year, the theoretical understanding of their effects on human behaviour is developed much more slowly.

The studies described in this thesis are intended to develop this theoretical understanding, specifically with regards to how we perceive and respond to emotion in such characters. This motivation stems from the intuition that there may be a future where social interactions with these characters may become the norm, and in this regard it becomes vital to understand how emotion communication and perception is achieved in the use of virtual characters. Additionally, with the advancement of these digital technologies, the need for narrative-based media (such as films, games and cartoons) to understand how empathic reactions are affected by increased visual and vocal realism in the presentation of these characters.

Aside from these motivations, more recent developments have seen a proliferation of animated virtual character technologies deployed in social media contexts. Given the dearth of content experienced by social media users and observed emotional reactivity associated with the personalisation of this content, the work of this thesis was motivated by the potential to understand the roles that social media based virtual characters can play in determining our emotional states.

A final important consideration is the use of virtual characters as representations of human speakers, in situations where real world speech is compromised. Virtual characters have been deployed for such purposes, meeting the needs of vulnerable populations as a means to communicate. This significant use of virtual characters provided motivation for this work to clarify how the presentation of vocal and visual features of these characters can be used to realistically portray emotions.

1.2 Scope and Limitations

The scope of this work, as with all doctoral theses, is subject to numerous limitations. The findings of this work cover niche touchpoints across multiple disciplines, and as such incorporate methods and models from the fields of cognitive science, human-computer interaction and computer graphics.

In brief, the scope of these touchpoints is one that may be reduced largely to screen based interactions with high fidelity virtual humans, based on principles drawn from embodied real-time animated scenarios. The link between these two experimental domains is one of explicit humanity; the humanlike characters used in these studies are in some sense responsible for the recognition of and response to emotional information. Additionally, since it has been established that virtual avatars in particular are designed in order to provide noticeable affordances for the purpose of interacting with virtual environments (for example, the mapping between an avatar arm and a VR controller), there is at the time of writing a significant gap in our understanding of the emotional affordances that such animated characters can provide for the purposes of social interaction.

This thesis thus takes its scope of assessing this recognition and response, in the perception of a virtual character, regardless of whether it is real-time animated and human-controlled, or preprogrammed to produce certain emotional expressions. The scope of this work is hence centered on the features of the characters which produce these perceptual activities in human users.

Outside of the limitations that will be highlighted later in this work, it must be stated from the outset that the scope of this thesis includes largely the study of virtual characters that resemble real humans, as opposed to virtual animals with different bodily affordances to humans, or social robots which are more explicitly depicted as non-human. One of the avatars employed in Chapter 3 is specifically designated as non-human, but retains the same humanoid bodily affordances (i.e. arms and legs and posture). The difference between this avatar and the humanlike characters used in this thesis is one of visual realism. Additionally, with regards to emotion, the scope of this work takes the literature on emotion seriously, and aims to situate research of virtual characters within the contemporary theoretical advancements of this domain. In doing so, this thesis focuses predominantly on the positive and negative aspects of emotional experiences, rather the degree of their intensity.

1.3 Thesis Overview

Within the defined scope, this work addresses empathic responses to realism in virtual characters. To reach this core topic, the work described in this thesis is divided as follows:

Background

The literature is used to explain the reasoning that brought us to reach our main research questions, namely: Do different degrees of vocal and visual realism influence experiences of empathy for virtual characters? To address this question, the background section is constructed with a view to defining and situating the terms used in this thesis. The first section of the background discusses definitions and applications of virtual characters. The second section narrows in on our operationalized definitions of realism, namely vocal and visual realism, with reference to the very broad scope that such a concept entails. The literature in this regard also covers the Uncanny Valley, where realism and emotion begin to interact in terms of virtual character perception. The third section builds from these concepts to situate the interpretation of emotion that was used as a framework for understanding our experimental results, specifically with reference to how emotion perception and empathy are 'afforded' to human beings via the perceptual and information processing mechanisms inherent to our cognitive architecture. The final section links all three of these sections together, and highlights the literature that makes scientific contributions which are similar to those of the present work.

The Subjective Experiences of Humanoid Avatars

This section describes our original formulation of research questions around experiences of human-controlled virtual avatars in a) emotive contexts and b) human vs nonhuman contexts. The results describe both a preference for positive valence emotions in virtual avatar interactions and a preference for human over non-human avatars.

The Effect of Unimodal Realism Variations on the Perception of Emotionally Expressive Virtual Humans

Following the previous work on subjectivity, human-likeness and emotional expressions, there appeared to be a gap in the literature specifying what exactly about these virtual characters, in their facial and vocal presentation, was influencing user perceptions. As such, these studies were designed to evaluate how variations in visual realism alone and vocal realism alone could impact emotion perception, self-reports of empathy and the

Uncanny Valley effect. Results describe only sparse significance in empathic response to unimodal realism variations. Additionally, more realistic characters were reported to be less Uncanny than lower realism counterparts, while emotion intensity was also considered significantly greater for high realism.

Empathic Responses to Bimodal Realism in Emotionally Expressive Virtual Humans using Facial Electromyography and Eye-Tracking Methods

The final experiment expanded on this variation in realism, to include both high and low realism conditions for appearance and voice. The inclusion of a real human voice let to greater reports of cognitive empathy, while evidence suggested that participants exhibited motor empathy for these characters regardless of realism. One interaction effect suggested that combined levels of high visual and vocal realism elicited stronger physiological responses.

1.4 Contributions

- The final experiment describes a novel methodology for measuring empathy in virtual character scenarios. This methodology may be employed in different formats, for Immersive VR or real human-human interaction contexts. This methodology, if replicated and validated in future works, serves as a pathway for understanding empathy at the physiological level in combination with self-report instruments.
- By conducting user studies in different contexts (video-conferencing, virtual reality, on-screen), this work contributes generalizable findings on how the presentation of virtual characters can influence human behaviour. These findings may be operationalised by developers in the gaming and film industries who seek to construct virtual characters to aid narrative presentation across different interaction modalities.
- Virtual characters are acquiring more relevance in the fields of medicine and education. The findings of this work can be understood to contribute important findings on emotion perception, empathy, social presence and realism preferences that must be considered for virtual character representations. These representations are used as tools for people with neurological disorders for the purpose of conversation, as well as training scenarios for doctors and other pedagogical settings. An important contribution of this work is to outline design recommendations for such systems.
- Significant findings on individual trait effects are also a notable contribution of this thesis. Users differ in terms of subjective traits; some people are more engaged with virtual characters, others less so. Some people are highly empathic, and others not so much. This work contributes novel findings on the nature of subjectivity in terms of how we as individuals relate to virtual characters. In the final experiment, this subjectivity is measured to have a novel physiological influence.
- Finally, we applied some measurements, specifically those used for assessing empathy, that had not previously been applied in the perception of virtual characters. The novel application of these tools serves to advance the literature on virtual character perception to explicitly include the notion of empathy, situated in the context of contemporary emotion research.

2 Background

In this chapter, the literature is evaluated in support of the core research questions. Given the multidisciplinary nature of these questions, the scope of the literature is considerably broad, and this is reflected in the diversity of content between the following sections. For the sake of clarity and readability, these are divided into four distinct areas.

In the first section, animated virtual characters are defined on clear terms. There is an outline of the techniques by which computer generated characters may be animated, while there is an emphasis placed on the difference between models of virtual characters, such as virtual humans, and embodied virtual characters, such as avatars. This section concludes with an analysis of research on both these kinds of virtual characters, as well as how the perception of these characters can be measured to be dependent on the individual traits of each user.

Secondly, the contemporary literature on emotion is discussed. In the context of this work, emotion is considered in terms of opportunities for social action provided by emotional cues in faces and voices, and the mechanisms by which these cues may afford empathy. This perspective incorporates numerous disciplines, and a great effort is made in this section to thoroughly define the terms on which emotion and empathy are considered for the experimental work that was conducted.

The third section is devoted to considering realism and how it is perceived. Like emotion, realism is a topic of remarkable depth, and is treated as such. While it begins with a broad scope, this is considered necessary for the context that it provides to findings on visual realism, vocal realism and their application in virtual characters.

Finally, the last section covers the different methods that have been employed in experimental settings to cover the relationship between emotion, empathy, realism and virtual characters. This wide ranging section is proposed as a summary of the different frameworks which motivated this thesis, in conjunction with the preceding literature on emotion and realism.

2.1 Animated Virtual Characters

Virtual characters, in the context of this work, are computer generated models of real people, that can be used as either substitutes for real people in simulated tasks or as self-representations within digital environments [Badler, 1997]. Such is the diversity of application for virtual characters, that it is necessary for the scope of this thesis to distinguish between particular subcategories of characters that have been evaluated in the literature:

- *Avatars*: Animated digital self-representations that a user may embody, manipulate and control during computer mediated communications [Nowak and Fox, 2018].
- *Intelligent Virtual Agents*: Animated virtual characters that can move and act independently to communicate with users in a manner which effectively simulates human intelligence [Aylett and Ballin, 2013].
- *Embodied Conversational Agents*: Computer generated characters that are designed specifically to simulate face-to-face conversation with users [Provoost et al., 2017].
- *Virtual Humans*: Human-like character models that can be used to exhibit human behaviours such as emotions, speech and gestures. Virtual humans may be used to represent all three of the categories above [Burden and Savin-Baden, 2019].

While avatars, agents and virtual humans may be represented in more or less human form, avatars are largely considered to differ in terms of degrees of control. In other words, an avatar is always human driven, and needs to be animated in real-time to match human movements. By contrast, agents are not necessarily animated in real-time.

Although virtual humans can be human driven and hence used as avatars, they may also be animated and controlled by instructions internal to a computer, in a manner which simulates agency. In this sense, virtual humans are the models which can be used to represent both agents and avatars [Fox et al., 2015]. As the work of this thesis progressed, it became clear that the representation of these models, without the element of control that defines an avatar, could be considered to be efficacious on the perception of emotion.

In order to properly convey emotion and other qualities, virtual characters need to be animated. This is to say that their movement in a given scenario requires visual depiction that allows for motion to be simulated in a virtual scene, through a sequence of changing images. In the field of computer graphics, there are numerous different ways a character

may be animated:

- *Hand-drawn animation*: Each frame of a sequence is drawn by hand. In modern animation, this 'hand drawing' is more often completed with computer-based drawing tools [Chen et al., 2005].
- *Procedural animation*: Frames are specified to represent different forms of motion by a given software technique. For example, key framing involves assigning values which correspond to changes in motion animation, while procedural scripts can also be written to change animated object properties over time [Carra et al., 2019].
- *Data-driven animation*: Changes in motion through time are specified by high level parameters, using data gathered from other animations or from motion-captured movement of real humans. The frames are generated using techniques such as deep learning which may predict the next animated frame based on a large quantity of animation data [Holden et al., 2016].

In producing animation with these techniques, a further distinction must be made between real-time and offline methods. While offline methods can take longer to produce convincing animations, the constraints on these animations largely stem from the applied technique, hardware and software capabilities of the animator. In real-time animation, there is a requirement to generate animations in a manner which is convincing but also quick enough to be rendered without artefacts. Hence, real-time animations are constrained by time in a way that offline animations are not [Weise et al., 2011].

Given that virtual human avatars are mapped to human movements, they are often required to be animated in real-time, and thus it is more difficult to generate convincingly realistic behaviours. In contrast, virtual human agents may be animated offline in a manner which allows for greater realism. Nevertheless, given that virtual humans may be used as both avatars and agents, this distinction is one where the separate categories often overlap. The focus of this thesis began with studies on avatar interaction, before zooming out to the evaluate specific aspects of realism in the character models used to represent humans in avatars. In this regard, the following sections deal with the definitions and applications of both avatars and virtual humans. Overall, the gaps that are depicted in the following sections relate to how we perceive and interact with avatars and virtual humans, and how these perceptions and interactions differ between subjects. This topic is addressed by our experiments, specifically with regards to how different kinds of people perceive embodiment and respond empathically to virtual characters.

2.1.1 Avatars and Embodiment

The word avatar stems from Hindu descriptions of descent into the physical world, and as such is applied today to the bodies produced by modern computer graphics systems which are necessarily embodied by human users [Ahn et al., 2011]. The term embodiment describes the way agents inhabit a physical body, particularly one that is embedded in an environmental context [Newen et al., 2018, Carney, 2020]. Embodiment may be considered as the core feature which allows for perception-action loops in living systems; in this sense, the adoption of a body allows us to perceive and act in the world. [Varela, 1997]. Embodiment can also be understood as a kind of coupling between an agent and its environment that enables cognition to take place [Ziemke, 2015]. Predictive processing (the method by which we interact with information by predicting possible outcomes in order to discriminate between percepts) approaches to modelling human behaviour propose agent embodiment to be an evolved method for minimising entropy (i.e. environmental uncertainty) [Chella et al., 2018], while there is strong evidence that the sensorimotor (how our senses relate to our physical actions) and spatial (where the body is situated in relation to other objects) contexts of a body are central to the performance of memory, attention and problem solving processes [Wilson, 2002]. In other words, bodies are what our minds are; we use them when we move, we use them when we perceive, and we use them when we think. Bodies are used for physical motion, but they are also used for cognition.

In Virtual Environments (VEs), the concept of *Afforded Embodiment* has been used to describe the relationship between a user and the functionality of an avatar body compared to a real body [Costa et al., 2013], with Gibsonian terminology (described later in this chapter) applied to emphasise the action opportunities that virtual avatars present to users. For the operation of humanoid avatars, these affordances are presented by computational algorithms which allow the human to drive the avatar as it is rendered in real time. Such algorithms control the appearance and animation of the virtual body, while also translating user actions into embodied agency, such as when the push of a controller button is mapped to the movement of a virtual hand in order to grab a virtual object [Bailenson and Blascovich, 2004].

From the user's perspective, the experience of this afforded relationship can be described in phenomenological terms as the Sense of Embodiment.The Sense of Embodiment (SoE) refers to the sense of being inside and in control of a virtual body [Fribourg et al., 2020]. Experiments conducted under the framework of embodied cognition suggest that users who embody virtual avatars tend to project their own bodily representations onto the avatar according to the affordances presented by the dimensions of the virtual body, in



Figure 2.1: Example of Avatar realism conditions from Fribourg et al and their experiment on point of view [Fribourg et al., 2020]

that their own preferences for bodily movement (e.g. being right or left handed) are carried through into virtual embodiment [Bailey et al., 2016]. These affordances can have different effects on user perception of an avatar. For example, both visual realism (see Figure 2.1) and the accuracy of virtually represented body movements have been shown to increase ownership and agency in virtual embodiment tasks [Argelaguet et al., 2016]. This influence of avatar realism is not specific to self-avatars, given that the realistic appearance of other avatars has been measured to heighten perceived self-avatar body change [Latoschik et al., 2017] in social avatar scenarios. These examples serve to highlight how our interactions with virtual characters and environments can give rise to illusory perceptions.

Illusions in Virtual Environments

It has been proposed that the illusions experienced in VE's and while embodying avatars are the result of sensorimotor contingencies [Gonzalez-Franco and Lanier, 2017] . That is to say, an illusion that occurs while embodying a virtual character can be considered a 'correction' to the predictions that are made by our minds when we engage in motor actions. Figure 2.2 serves to describe this model of sensorimotor contingencies, and relies heavily on the notion of 'top-down mechanisms' for the experience of these illusions. This means that our treatment of virtual objects as real objects is considered to be reliant on compiled sensory information that is corrected against predictions that we learn to make in the real world.

However, there remain open questions as to the full nature of this mechanism, and specifically whether it occurs in the same way for object interaction as it does for social interaction. The malleability of self-identification has emerged as a key issue for addressing such questions. While SoE serves as a general term for the illusion of bodily representation in avatar use, it can be described in terms of several sub-components; agency, body ownership and self-location [Kilteni et al., 2012]. Agency is the degree to which participants feel in control of a virtual body, ownership denotes the feeling of the body belonging to a user, while locatedness captures the sense of being positioned inside

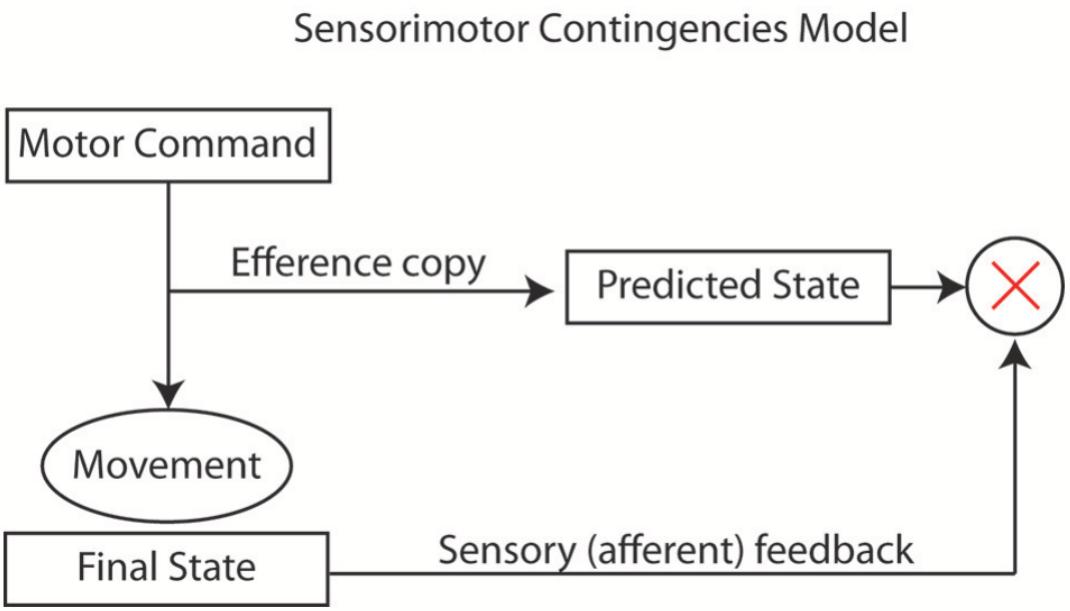


Figure 2.2: The sensorimotor contingencies model proposed by [Gonzalez-Franco and Lanier, 2017] which is alleged to describe the mechanisms by which virtual reality illusions take place.

a body [Kilteni et al., 2012].

These dimensions have each been targeted in empirical works. For example, studies focusing specifically on agency [Jeunet et al., 2018] have used electroencephalography (EEG) methods to identify neurophysiological markers of both the feeling and judgement of agency in virtual embodiment. Similarly, studies on the sense of ownership [Kokkinara and Slater, 2014] have highlighted the influence of simultaneous visual and tactile or visual and motor stimulation as important for inducing body ownership.

Senses of body ownership can be understood as a fusion of place and plausibility illusions [Slater, 2009a] where a user attributes their body to themselves as if the body were the source of experienced sensations [**?**]. Place and plausibility illusions can be considered extensions of the same concept, whereby a user reacts to virtual stimuli as if they are real, with place referring to the realism of a place, and plausibility more generally referring to the sense of plausibility in a given virtual scenario [Slater, 2009a]. Similarly, this can extend to the feeling that a non-human virtual character exhibits a form of social presence without an actual human being present [Biocca, 1997] inviting the illusory perception of a 'real' other in a virtual environment [Biocca et al., 2001, C.S. Oh and Welch, 2018]. This is an aspect of presence distinct from the positional notion of telepresence which mediates modelling of spatial environments in virtual contexts [Biocca et al., 2001], and which has links to illusory sensations of embodied placement [Slater, 2009b]. Partic-

ularly, maintaining personal distance from a character, similar to real life encounters with people, reveals user's comfort with the character and can be observed and measured in VR to influence social presence [Bailenson et al., 2003]. While previous studies investigating social presence with avatars have failed to define a relationship to either agency or body ownership [Nowak and Biocca, 2003, C.S. Oh and Welch, 2018] recent work has implicated personalisation (See Figure 2.3)of avatars as a potential modulator of experiential presence and virtual body ownership [Waltemate et al., 2018]. Similarly, Embodied Social Presence theory posits SoE as a necessary precondition to experiencing any categories of presence in virtual worlds, with bodily experience considered as the primary bottom-up sensory modality required for the physical and social actions contingent on presence [Mennecke et al., 2010].



Figure 2.3: Personalised avatars used in the experimental work of Waltemate et al [Waltemate et al., 2018]

The term 'enfacement' has also been coined as a means of defining the plasticity of subjective senses of face ownership, which can be adapted according to features of virtual faces [Sforza et al., 2010, Porciello et al., 2018]. In an analogous way to the SoE, the enfacement illusion can be considered as a model instance of self-identification and involve the same kind of multisensory integration (i.e. combination of different kinds of sensory information, such as in visuotactile stimulation) applied to the face [Tajadura-Jiménez et al., 2012]. Enfacement is an extension of embodiment, specific to adopting the movement of a virtual face, and has measured to conform to similar constraints in experimental scenarios. For example, animation quality has been a factor documented to influence levels of reported enfacement with avatars [Gonzalez-Franco et al., 2020], while there has also been previous research linking facial mimicry of others to strong senses of face-ownership [Minio-Paluello et al., 2020].

A similar perceptual illusion that can occur in virtual embodiment is known as the

Proteus Effect. This phenomenon is understood as a form of behaviour change that stems from avatar embodiment [Yee and Bailenson, 2007], where the presentation of an avatar body can be measured to alter perceptions of self external to the use of the avatar [Heide et al., 2013]. For example, users embodied in attractive avatars reportedly consider themselve to be friendlier compared to less attractive avatars [Yee and Bailenson, 2009], while avatars presented under negatively valenced conditions can also provoke more aggressive actions from users [Peña et al., 2009a].

Avatars in Research

Avatar use in VEs is regarded as a fruitful platform for social neuroscience research [Parsons et al., 2017], facilitated by the dirigible qualities of virtual environments, which enable researchers to control the audio-visual and tactile scene data that experimental subjects experience, to the most minute of details. Studies utilizing these advantages have been able to examine non-verbal social behaviours and cues [Roth et al., 2018] as well as task-coordination, joint attention and eye gaze in digitalised social scenarios [Roth et al., 2018, Wienrich et al., 2018]. Work in this domain has also sought to quantify the effects of smile enhancements [Bailenson et al., 2016] and pro-social intentions [Roth et al., 2019] (See Figure 2.4).

Studies have verified that mimicked face and lip movement are sufficient to induce a sense of self-identification with an avatar, even when a face is not directly mimicking those of the user [Gonzalez-Franco et al., 2020], which can be used in virtual environments where head mounted displays prevent accurate facial mapping. The effect of enfacement can be shown even if the animations are pre-generated. However, measures for self-identifying with an avatar do scale up once lip-syncing is enabled and facial motion is mapped to the user. In carrying out the experimental work of Chapter 3, we sought correlations between this assimilation and the notion of enfacement [Gonzalez-Franco et al., 2020].

Avatar studies have also empirically validated self-avatar (avatars that are made to look like realistic representations of an individual user) influences on cognitive load [Steed et al., 2016], pain modulation [Romano et al., 2014] and perceptions of self-concept [Yee and Bailenson, 2007, Heide et al., 2013]. These specific applications of avatar-interaction research have been developed from a volume of work which has attempted to define the qualitative feeling of being a virtual representation of oneself.

In sum, the literature on avatars and embodiment informed the experiments for Chapter 2 of this work, which focus on the sense of embodiment in virtual avatars. However, as will be explicated, the visual and vocal affordances inherent to these experiences encouraged us to investigate how they might afford different opportunities when applied

to independent virtual humans that operate outside human control. Additionally, there was a gap in the literature with regards to treating both avatars and virtual humans using the same framework of afforded opportunities for action.

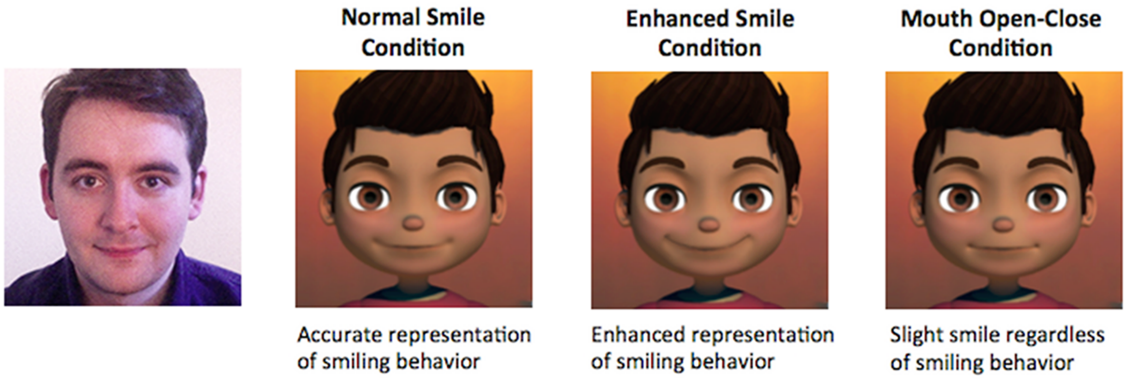


Figure 2.4: Face only avatars used to by Oh et al to test the effects of enhanced smiling behaviour [Bailenson et al., 2016]

2.1.2 Virtual Humans

Avatars, as outlined above, are typically animated depictions of bodies, designed and rendered in accordance with certain movements on behalf of a user.

Virtual humans, on the other hand, are not always human controlled. As such they can be animated with a set of predetermined movements, actions and expressions that are dictated by a given circumstance (e.g. initiation of contact by a user), but can also retain their own agency and perceived freedom of movement. While, particularly in gaming contexts, virtual humans can be influenced by users, they are importantly not always controlled by them, and as such do not afford embodiment when they come in the form of independent agents. As this thesis progressed, the focus graduated from considering the well documented affordances of virtual avatars for embodiment, towards the aspects of the character representations themselves (without an internal locus of agency attached) that influence our responses to their expressions of emotion (See Figure 2.5).

Although they may be used as avatars, virtual humans are often used in a variety of other applications as independent characters [Ahn et al., 2011, Rekik et al., 2024]. A virtual human is a more complex concept than an avatar; while virtual humans may be simple bodies expressing human-like qualities such as appearance, voice, emotional qualities (which were the main focuses in this study) they can also be designed with more advanced techniques, in order to produce more human-like behaviours such as senses, reaction triggers, language capabilities and the ability to create and pursue goals, which add additional layers of complexity to their potential functions and in-

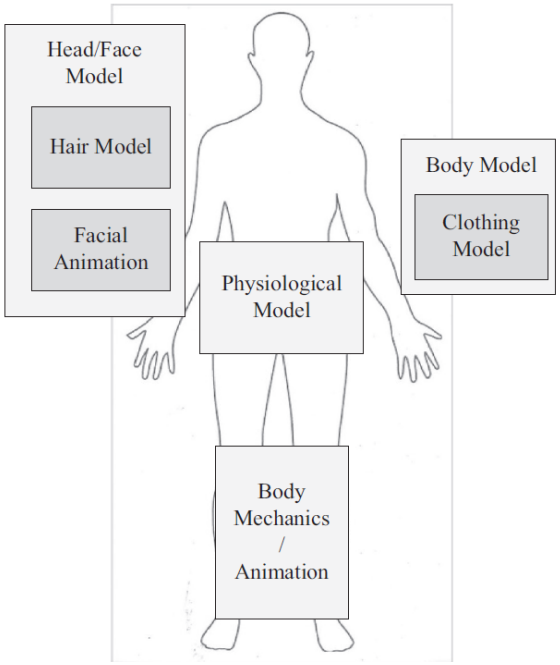


Figure 2.5: Components of a virtual human body [Burden and Savin-Baden, 2019]

fluences on perception [Burden and Savin-Baden, 2019]. The decision to focus on the appearance, voice and emotion aspects of virtual humans was motivated by the the fact that voice and emotion are the primary factors in users attributing agency to a system [Reader and Savin-Baden, 2020], and are thus the main factors which separate a virtual human from an avatar, in that the locus of agency is shifted. In addition, the focus on visual appearance was motivated by similar theoretical notions that have stemmed from the field of robotics, with regards to appearance realism and emotional responses [Mori et al., 2012]. The central questions to be explored on the topic of virtual humans have been developed from an evolving foundation, where technological advancements have improved our ability to faithfully simulate human appearances, voices and emotions. Specifically, the aspect of facial appearance, as highlighted in this study, has been the focus of significant technological advancements in recent years. In order to generate realistic virtual faces that are capable of expressing emotions, developers are often required to create an detailed 3D model that can be used to animate believable movements, as well as texture details and spectral maps which maintain facial details such as wrinkles across changes in illumination that are typically caused by face movements [Burden and Savin-Baden, 2019].

Understanding interactions with realistic virtual humans has become increasingly relevant, as methods for generating veridical character models continue to develop at an impressive rate. Contemporary studies have sought to develop our understanding of dif-

ferent psychological effects in different contexts, such as how virtual humans can simulate patient interactions in medical scenarios [Kleinsmith et al., 2015, Halan et al., 2015], as well as developing our understanding speech processing and intelligibility in lab scenarios [Devesse et al., 2018, Thézé et al., 2020] and trust in cooperation games [Torre et al., 2018] to name but a few.

As highlighted in the later section on *Affordances*, there are several parallels to be drawn between the way we treat people and the way we treat computer generated characters. In particular, virtual humans may be considered human-like based on a number of different properties which encourage users to treat them as real social agents; while there are properties such as appearance, voice and motion which can be considered as cues to human animacy, users are also influenced by prior beliefs. As will be discussed in the next section, this can include preexisting traits, as well as information acquired prior to an exposure to a virtual human model [Cross et al., 2015]. This is especially relevant to the present thesis, given that our prior beliefs and known concepts are not only important for our perception of virtual characters as social agents, but also contribute to our perception of emotion. Moreover, the aforementioned concept of social presence can also play a role in our social perception of virtual human models, as demonstrated by work which has found similarities in how we perceive intimacy with virtual and real human advisors, in that a sense of social presence is tied to a sense of intimacy [Potdevin et al., 2020].

Numerous aspects of virtual humans can increase this sense of social presence. There is some indication that the self-reported social presence is higher when observing visually realistic characters as opposed to more stylised rendered virtual humans in virtual reality [Zibrek and McDonnell, 2019, Zibrek et al., 2017], and can also be increased when the appearance of the character matches its behavior [Zibrek et al., 2018]. There is also evidence showing the importance of haptics for increasing social presence [Sallnäs, 2010], while a study investigating audio in virtual reality found a positive relationship between audio quality and the sense of social presence [Skalski and Whitbred, 2010]. While there are many other determinants of social presence [Oh et al., 2018], not much is known about the importance of quality and naturalness of the character's voice, although there is evidence that the close resemblance of a virtual human's voice to one's own voice can lead to greater senses of immersion and stronger task performance in virtual gaming environments [Kao et al., 2021].

Moreover, it has been proposed that mismatching human-like conversation qualities can negatively sway perceived social presence in interactions with virtual humans [Hu et al., 2019]. Social presence has been measured to increase in experiments using synthetic voices where verbal utterances are matched deliberately with correspond-



Figure 2.6: Virtual humans presented at different degrees of visual realism, as employed by Parmar et al [Parmar et al., 2022]

ing personality cues [Lee and Nass, 2003] suggesting that social presence measurements could be altered by a misalignment of visual and vocal realism.

As the focus of this thesis progressed, the literature revealed that this sense of social presence could allow for opportunities to measure the treatment of virtual humans as social agents rather than animated objects. It has previously been shown that our reactions to both static and dynamic images of emotionally expressive virtual humans can evoke similar responses to static and dynamic images of real humans [de Borst and de Gelder, 2015]. As explored in this thesis, manipulating different aspects of virtual humans has been documented to change user perceptions of those characters, including voice and appearance (as seen in Figure 2.6) [Parmar et al., 2022]. However, there is also a subjective element to these perceptions, which can be considered from both 'bottom-up' stimulus perceptual perspectives (i.e. the way an individual user perceives attributes of a virtual character) as well as top-down knowledge cues (i.e. the user is primed in a certain way, or believes certain beliefs, that change how the virtual character is perceived) [Cross et al., 2015].

2.1.3 Subjective Experiences of Virtual Characters

From the sections above, it is clear that virtual characters can have numerous different effects on human behaviour. In other words, the perception of these characters can

be dictated by the presentation of the characters, whether they are avatars or agents. However, there is also a considerable amount of work on the antithetical relationship; what about an individual subject, their exposure to previous stimuli or their personal traits and experiences, influences the perception of these characters? In this section, we focus on the subjectivity of these experiences, before we analyse affordances, realism and emotion in the following sections.

Afforded embodiment in VR allows for opportunities to test how stimuli presented prior to avatar exposure might influence the Sense of Embodiment. Researchers' uses of priming methods have been intended to capture the wide range of content that may influence perceptions of virtual avatars. Concept priming may be understood as the controlled introduction of a particular concept without a subject being explicitly aware of its influencen [Bargh and Chartrand, 2000]. The term priming, as it is used today, has been attributed to linguistics research from the middle of the twentieth century [Lashley et al., 1951]. Priming studies in psychology began in earnest after work that described the effects of semantic context priming on reaction times for lexical decision making tasks [Meyer and Schvaneveldt, 1971].

Although some research in this area has been criticized on replicability grounds [Weingarten et al., 2015], there have been efforts to extend this work further in the domain of virtual characters, largely in immersive virtual reality, with some of these outlined below. Since VR necessarily includes an interactive environment with provisions for different sense modalities, it is fitting that contemporary models of priming effects describe their underlying processes as inter-sensory, interactive heuristics based on perceptual associations [Klatzky and Creswell, 2014].

Most research on priming and VR has focused on how avatars can prime subjects for different behaviours. This priming influence has been termed The Proteus Effect, whereby levels of avatar attractiveness are proposed to increase levels of participant friendliness (See Figure 2.7) [Yee and Bailenson, 2007] which describes the behaviours that avatars can prime people to exhibit during or post-embodiment [Peña et al., 2009b]. Avatar appearance and behaviour have since been associated with, among other things, changes in walking speed [Reinhard et al., 2020], self-objectification [Fox et al., 2013] and creative ideation [de Rooij et al., 2017]. Research in this area has proposed the effects of particular avatar cues on perceptual associations to be reliant on automaticity, mediated by activation and inhibition spreading mechanisms, as well as valence processes that are intrinsic to human memory function [Peña, 2011, Bargh and Chartrand, 2000].

While the efforts to investigate avatar-human priming effects have been fruitful, less work has been done at the intersection of priming influences prior to VR exposure.



Figure 2.7: Avatars at different levels of attractiveness have been shown to influence user behaviour [Yee and Bailenson, 2007]

Some work has found an effect of instructional priming on subjects' movement within virtual crowds, exhibiting greater collision avoidance behaviour when instructed that a real person's movement was represented by the crowd [Gupta et al., 2017]. Similarly, haptic priming has been implicated in the improvement of social responses to virtual humans [Krum et al., 2018], while priming for meditation and positivity has also been shown to reduce student anxiety [Hawes and Arya, 2021]. Aforementioned work has proven that our method of concept priming can increase the sense of presence in VR [Nunez and Blake, 2003, Cerda et al., 2021]. While these studies focused on activation mechanisms developed from interacting with previous environments, other studies have chosen to prime participants with introductory concepts for improving experiences in VR. Presence in virtual environments has been shown to improve with the use of a visual preamble that mirrors their own environment [Smolentsev et al., 2017], while introductory limbo states can heighten perceptions of virtual body ownership [Jung et al., 2018]. Additionally, mindfulness priming can enhance responses to neurofeedback training [da Costa et al., 2021], and has also been shown to negatively impact responses to the rubber hand illusion [Cebolla et al., 2016], and to reduce stress responses to high arousal VR situations [Crescentini et al., 2016].

Measurements of Subjective Traits

While we were interested in examining whether subtle differences in external cues, such as a primed concept or the presentation of a virtual character, could be measured to effect the perception of virtual characters, the literature also presented an opportunity to expand previous findings on the role of internal cues, such as personality traits, specifically in terms of the Five Factor Model, as well as trait empathy, measured by the Interpersonal Reactivity Index, as well as perceptions of body awareness.

Body awareness can be defined as the perception of bodily states, processes and actions, based on proprioceptive and interoceptive sense data within the awareness of an individual [Mehling et al., 2009]. Some evidence suggests that trained mindfulness meditators can exhibit higher body awareness [Treves et al.,], but some objective measures of internal bodily awareness, also termed interoception, do not support this link [Silvanto et al., 2015]. Given this issue of understanding how mindfulness influences VR experiences, the present study included a measure of body mindfulness extracted from the State Mindfulness Scale [Tanay and Bernstein, 2013], which has previously been employed in Virtual Reality settings [Cebolla et al., 2016, Döllinger et al., 2022] to assess aspects of body awareness.

It has been demonstrated that self-reports of interoceptive body awareness correlate with heartbeat perception scores in VR tasks [Döllinger et al., 2022]. Participants who exhibit lower body-awareness on self-perceived heart rate measurements also experience less ownership over a rubber hand [Tsakiris et al., 2011]. While there is some evidence that body awareness is not disrupted by VR body-swap tasks [Cebolla et al., 2019], mindfulness meditators have reported lower agency during the rubber hand illusion [Cebolla et al., 2016]. The volitional aspect of self-regulation, measured in our study, has also been shown to benefit from, as well as support, mindfulness practice, with reports indicating greater agency in terms of both executive control [Teper and Inzlicht, 2013] and motor control [Naranjo and Schmidt, 2012] for mindfulness meditators. Uncertainty around the influence of interoceptive body awareness on VR experiences demonstrated a gap in the literature which the study outlined here aimed to partially fill, through the use of the Multidimensional Assessment of Interoceptive Awareness (MAIA) [Mehling et al., 2012].

As well as individual differences in body awareness, individual differences in personality traits have been highlighted in the literature as potentially influential over experiences of virtual reality [Sas and O'Hare, 2003] and embodiment during the rubber hand illusion [Burin et al., 2019, Lush et al., 2020]. A common theoretical perspective on personality traits are to be found in the Five Factor Model [John et al., 1999, Gosling et al., 2003];



Figure 2.8: Dewez et al 2019 assessed the relationship between Five Factor personality traits and virtual embodiment

a set of measured and widely replicated traits that are used in efforts to taxonomise the plurality of human personas, namely Extroversion, Agreeableness, Conscientiousness, Emotional Stability and Openness [Cobb-Clark and Schurer, 2012]. These traits have been studied relative to spatial presence [Sacau et al., 2008] and immersion in virtual reality [Weibel et al.,], as well as embodiment [Dewez et al., 2019a] (See Figure 2.8). Furthermore, the dimensions of neuroticism and agreeableness have been modelled computationally for the purpose of providing personality traits to embodied conversational agents [Bevacqua et al., 2012].

In essence, the work above was used to establish the potential influence of subjectivity on the experience of empathy for virtual characters. While much of this work focuses on avatar embodiment, it also allowed us to narrow in on specific traits which may influence perception of characters in dyadic scenarios. While the personality and body awareness traits mentioned above proved beneficial to the domain of human-avatar studies, our other findings hinted that it was the nature of the presented character itself that was influencing responses.

Nonetheless, the same Five Factor Model has been applied in assessing various changes in perception of independent virtual humans, such as how appealing a virtual human is to an individual user [Kang et al., 2008], how they generate rapport [Von Der Pütten et al., 2010], and whether they are considered trustworthy [Zhou et al., 2019]. Moreover, other studies on interactions with virtual humans have evaluated how subjective levels of social anxiety can influence perception of realism in virtual humans [Kwon et al., 2009] as well as social presence [Garau et al., 2005]. Additional individual factors which are documented as impactful on our interactions with virtual humans include senses of self-efficacy [Zalake et al., 2024], speaking confidence [Loveys et al., 2020] and social concern [Qi et al., 2021].

One of the subjective characteristics measured in the treatment of virtual char-

acters in general is trait empathy, also referred to as dispositional empathy [Sierra Rativa et al., 2020]. As such, in the final experiments, we employed the Interpersonal Reactivity Index to assess this potential influence. The Interpersonal Reactivity Index [Davis et al., 1980] is a commonly used survey instrument that contains four separate subscales measuring different dimensions of trait empathy (Empathic Concern, Perspective Taking, Personal Distress and Fantasising). For instance, the Empathic Concern subscale may be considered a stable measurement of trait emotional empathy, while the Perspective Taking subscale can be understood as trait cognitive empathy [Cowan et al., 2014]. While there are contentions that Personal Distress cannot be considered a standalone measure of empathy [Pulos et al., 2004], it was included in our analysis in order to capture any possible effects. This measurement was to broaden our empathy focus. While the questionnaire used in at the end of each block in experiments 2C and 3 from [Powell and Roberts, 2017] was intended to capture state empathy, at three levels of Cognitive, Affective and Compassionate empathy for each character condition, it was important to contextualize any such reports by recording empathic traits that individuals displayed to greater or lesser degrees. Of particular interest were the Perspective Taking, Fantasising and Empathic Concern scales, given their associations with Cognitive and Affective empathy [Jolien Van der Graaff and Branje, 2016]. In general, this instrument has been used for assessing a variety of user responses to virtual character scenarios [Rodrigues et al., 2015, Gao et al., 2017].

In sum, the use of virtual humans and avatars is widespread, and as such can touch off an individual's perceptual schema in numerous different ways. Many of these differences can be captured in terms of subjectivity or primed associations. However, as our studies also show, much of this influence can also be reduced to the influence of realism, as well as the influence of emotional affordances inherent to the presentations of such characters. The following two sections are written to reconcile contemporary trends in the sciences of emotion and realism with the study of virtual characters.

2.2 Emotional Affordances

The origin of the word emotion comes from the latin word for move (mouvere), and as such has been associated with action, rather than experience or feeling, since its very conception [Neta and Haas, 2019]. This is relevant to the current thesis, which considers emotion from an enactive framework, as a constructive action which relies upon particular affordances which are perceived by a person as specific to the actions of a virtual character. Emotional affordances, in this context, are any mechanisms by which emotional information is contained or transmitted by an object, that allow perceivers to construct an instance of an emotion to guide further action [Vallverdú and Trovato, 2016]. In the facial and vocal expressions of virtual characters, it may be said that emotional affordances provide cues for users to label an emotional state, as a form of heuristic for interpreting beliefs or intentions of an emotional character, and hence serve as a guide for a user's next action [Poggi and Pelachaud, 2001, de Melo et al., 2023].

2.2.1 Models of Emotion

In general, emotion research has a long and debated history. This history is littered with errors and mistaken assumptions, so that many interesting questions still remain for researchers to address. Famous philosophical figures have attempted to address these questions. Aristotle once defined emotion as "that which leads one's condition to become so transformed that his judgment is affected, and which is accompanied by pleasure and pain". Similarly, Descartes referred to emotions as "perceptions of the soul" [Solomon, 1999].

The first statement is perhaps overly reductive, while the second is certainly more than a little melodramatic. However, they both hit on something true about emotion; namely that it is an essential feature of everyday experiences. Emotion is not just thought, it is not just experience and it is not just memory. The prevailing intuition is that emotion seems to to be a necessary part of all of these phenomena, but yet none describes it. The current thesis considers emotion to be one of the main perceptual mechanisms that allow us to perceive, think and act.

Emotion is often split into the dimensions of valence and arousal. This split is commonly associated with the 1896 work of Wilhelm Wundt, who reduced all psychic elements to 'sensations'. He was recorded to have described these sensations in detail, as an account of conscious experience and in an effort to establish psychology as a science. His main contributions seem to be two particular attributes of sensations; affective tone, which is categorised as pleasant or unpleasant, and intensity, which may be high or low. These

have since been developed and understood as valence and arousal in emotion research [Reisenzein, 2000].

Emotion perception may be understood as the identification of emotionally salient information that can be used as cues for understanding the emotions of others [Mitchell and Phillips, 2015]. This information can come in many forms, from lexico-semantic cues, to body movements and appearance. However, valence and arousal themselves have been posited as a source of embodied emotional information that can prime for the anticipations of actions, and hence alter our perception by themselves in numerous ways [Zadra and Clore, 2011].

In discussing emotion, it is necessary to caveat any scientific work with some important clarifications on the topic. A 2010 study [Izard, 2010] sought to improve the operational capacity of the term *Emotion* by asking 37 prominent research specialists to define precisely what it is, its unconscious functions, its regulation, the causes of its activation and the most pressing questions around it. This study was qualitative in method, and returned a grand sum of 34 entirely different definitions. This is not an altogether original finding. It has previously been observed by Fehr and Russell [Fehr and Russell, 1984, Beck, 2008] that "Everyone knows what an emotion is, until asked to give a definition. Then, nobody knows". As such, emotions and their expression are the subject of great debate.

As noted by [Kleinginna and Kleinginna, 1981], this ambiguity is in keeping with older efforts to define this universally understood yet uniquely vague concept, which has previously been discussed in the contexts of cognitive experiences, affective phenomenology, behaviourism, expressive behaviour, adaptive behaviour, physiological responses, and, more recently, constructive perception [Barrett, 2011, Barrett et al., 2019]. Constructive perception is at odds with more materialistic approaches to the study of emotion, which hold that our experiences of emotion can be reduced to the functions of innate brain circuits [LeDoux, 2012].

It is this notion of constructive emotion perception [Barrett, 2011, Barrett et al., 2019] which is relevant to the forthcoming subsections which cover emotion perception through the frameworks provided by 'direct perception', whereby it is necessary to describe objects, including virtual characters, their faces and their voices, in terms of affordances. Theories of constructed emotion have recently provided insight into the ubiquitous subjectivity and active perception of emotion [Barrett et al., 2019]. This approach contrasts with traditional conceptions of emotion expression, exemplified by the theory of 'basic emotions', which highlights a discrete set of emotions that are expressed in more or less the same way by all humans, through facial expressions, or vocal expressions, or

other singular representations of an emotion [Ekman, 1999]. These basic emotions have been operationalised for the development of the Facial Action Coding System (FACS); computationalised movements of real face muscles which have formed the basis of animating virtual faces in the majority of modern systems for generating virtual characters (See Figure 2.9) [Tolba et al., 2018, Seymour, 2019], including the characters used in the present thesis. In the facial action coding system, there are 80 different 'action units' which are used to portray and objectively measure emotional expressions from particular presentations of the face, head, eyes and miscellaneous movements such as tongue protrusions [Prince et al., 2015] which are broken down into upper and lower face movements. One example is the combination of two eyebrow movement units, whereby the brow a) pulls together and b) pulls upwards, which is proposed to represent sadness [Cohn et al., 2007].

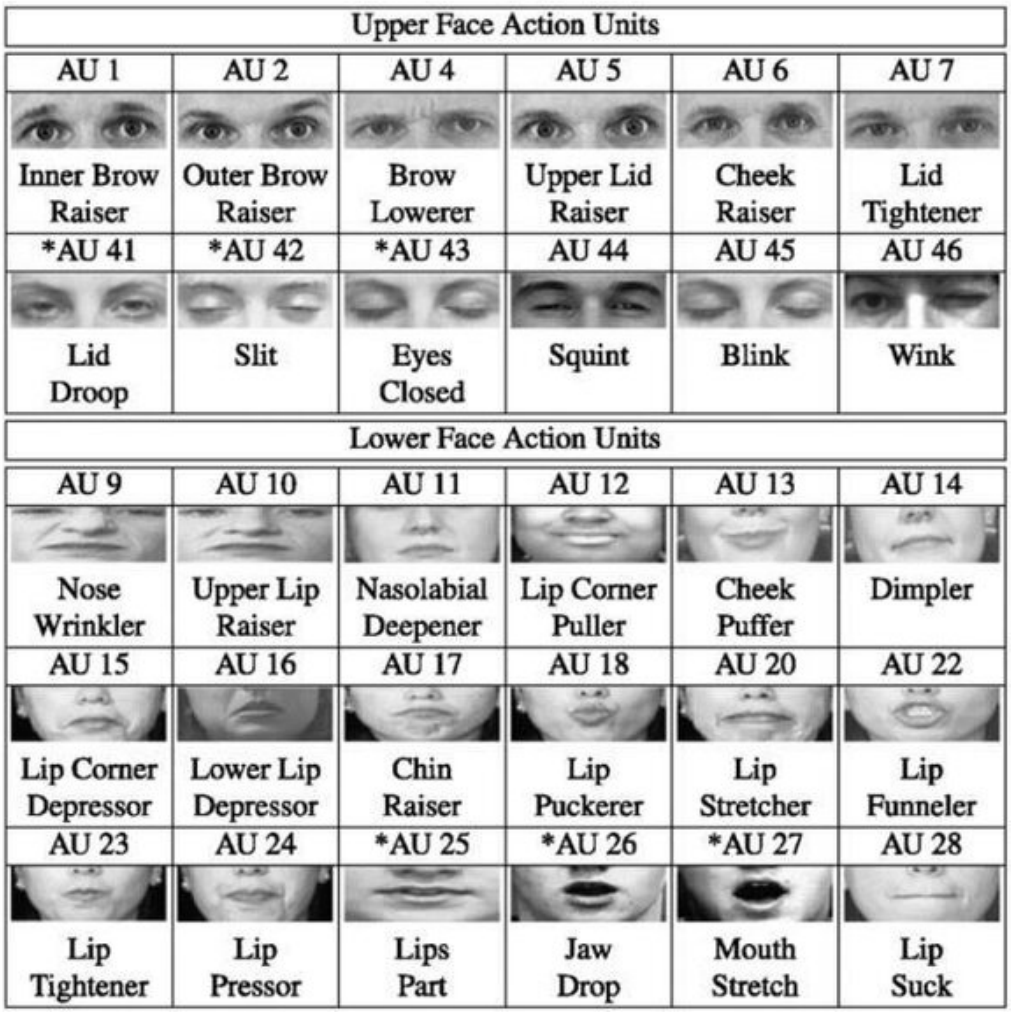


Figure 2.9: The facial action coding system applies numeric values to muscle movements in the human face that are used for facial expressions of emotion [Hassouneh et al., 2020]

In practice, it seems obvious that there arises a mutual influence of constructed emotions, learned concepts and actionable affordances for emotions in animated faces. Interestingly, it is this constraint of afforded cues that differs in our regular perception of emotions; in real people, there are numerous ways to express or perceive different emotions, that are not shared by every person. However, with virtual characters, they are more often designed on the assumptions of basic emotion theory, powered by the FACS, and as such represent the way we typically see emotions expressed in animated faces. In this regard, it seems that the universal basic emotion expressions afford the emotions that they are contended to afford. That is of course, until they are combined with realistic or unrealistic voices, not to mention other affective human-like features. At that point, the necessity for a constructive approach to emotion perception takes primacy. The basic emotion perspective and the constructivist perspective are two models of emotion that differ on the basis of constraints. While the basic model considers constraints to be inherent to the expression (e.g., we are constrained to perceive a frown or a smile and hence interpret those emotions as positive or negative respectively) constructivism is modelled on cultural and contextual constraints (for example, a smile can be threatening or malevolent, while a frown can be acted for comedic affect). These constraints are the result of learned concepts, rather than universal fingerprints or neural representations, that enable us to perceive afforded emotions in the expressions of both real and virtual characters [Barrett, 2017].

In this thesis, we employ the two dimensions of valence (positive and negative) in order to measure how differing emotional expressions in virtual characters may influence empathy in users. While this dimensional approach is a very common feature of most emotion models [Sander, 2013], there is a necessity to justify the use of particular discrete emotions (specifically happiness and sadness) as synonymous concepts. To this end, it may indeed also appear obvious that these represent 'basic' emotions rather than constructed ones. However, this thesis would follow other works in contending that discrete emotional states are influenced by individual's attitudes towards those states; in other words, the positivity of happiness and the negativity of sadness are themselves subject to individual assessment, even if their dimensional certainty as positive and negative is generally accepted by most people [Harmon-Jones et al., 2017]. As such, this work considers discrete emotional states on the scale of possible dimensions of positive and negative valence emotions to be interpretable in a broad range of ways depending on the individual perceiver and their attention to different emotional cues. Moreover, it is shown in Chapter 5 that vocal characteristics in virtual faces appear to define the scope of these cues. That is to say, while an animated face, powered by representations of basic emotions, can afford emotion perception, this task of emotion perception becomes

much easier with the presence of a voice, and easier still with the presence of a real human voice. It is the compilation of these cues together that affords emotion.

While it can be very obvious from observing a facial expression as to what emotion an agent is expressing, a helpful analogy may be described by the language of affordances (which are explained in detail in the next section). For example, a door-handle, placed at the center of a door, is not only a piece of metal jutting out from a piece of wood, it is also a concept that we as human beings learn to associate with the opening of a door. In other words, when we perceive a door-handle, we perceive its affordance for opening a door. However, that door-handle, although it can be used for opening the door, does not tell you if the door is locked, unless you try to open it, but even then one cannot tell whether you should push or pull on the door, where the door goes, or other relevant information as to its opening. While the affordance for opening the door exists with the perception of the door-handle, it is not helped by any other medium apart from the concept of a door-handle affording the door's opening. However, on many doors that are used in public spaces, there are often signs, words or semantic cues attached to the door; a sign that will say no entry, or exit only, what times the door is open, why it is locked, or whether the door should be pushed or pulled in order to be opened. It is not that one cannot figure out, from just the door-handle, how to open the door. But it is much more ambiguous without the signage, and hence it requires less mental effort to open the door if there is a sign present. It is contended in this thesis that, following the results of our final study, this analogy is a useful one for describing our perception of emotion in faces and voices. In essence, when it comes to understanding an expressed emotion, a virtual facial expression is analogous to a door-handle. A synthetic voice attached to that facial expression, expressing the same emotion, is similar to a blurry sign underneath the door-handle, that dictates whether it should be pushed or pulled in order to open it. And a real voice, attached to that same virtual facial expression, expressing that same emotion, is akin to a bright neon sign that says push, the opening hours for the door, and where the door leads. This compilation of cues is an additional affordance to the door-handle or the sign by itself. Similarly, the stacked concepts of a facial expression and a vocal expression are themselves a perceived affordance for the same thing as one or the other alone. It is just a clearer affordance for emotion when the cues are combined, an affordance that in a sense is more useful and contains more information than its constituent parts.

In contending that emotions are perceived through a process of compiling different cues, it becomes necessary to adopt that vocabulary of affordances, which more generally includes the assumption of perception as an activity which in itself is for generating further action.

2.2.2 Affordances

The term 'affordance' originates from the work of JJ Gibson, and has been utilised to study the perception of virtual environments and characters [Morie et al., 2005, Van Vugt et al., 2006]. Gibson's 1966 opus *The Senses Considered as Perceptual Systems* [Gibson, 1966] popularised the use of the term affordances for understanding perceptual behaviour. Originally, an affordance was conceptualised as a non-numeric value inherent in a given object:

> **When the constant properties of constant objects are perceived (the shape, size, color, texture, composition, motion, animation, and position relative to other objects), the observer can go on to detect their affordances. I have coined this word as a substitute for values, a term which carries an old burden of philosophical meaning. I mean simply what things furnish, for good or ill. What they afford the observer, after all, depends on their properties. The simplest affordances, as food, for example, or as a predatory enemy, may well be detected without learning by the young of some animals, but in general learning is all-important for this kind of perception. The child learns what things are manipulable and how they can be manipulated, what things are hurtful, what things are edible, what things can be put together with other things or put inside other things - and so on without limit. He also learns what objects can be used as the means to obtain a goal, or to make other desirable objects, or to make people do what he wants them to do. In short, the human observer learns to detect what have been called the values or meanings of things, perceiving their distinctive features, putting them into categories and subcategories, noticing their similarities and differences and even studying them for their own sakes, apart from learning what to do about them.** (Gibson, 1966 p. 285 [Gibson, 1966]).

This portrayal of sensation, as an interaction between perceptual systems that dynamically capture information inherent to space and objects, has proven to be incredibly influential in framing contemporary views of the mind non-computationally, as necessarily embedded, extended, enactive and embodied [Carney, 2020]. Its most relevant contribution to the present work is this split between perception and sensation (See Figure 2.10). On Gibson's terms, perception is considered to be an activity of adapting passive sensations to produce information about a given situated environment. As such,

perception is not as simple as 'stimulus exposure'. Perception is a task that never ends, which involves objects of many different kinds, each with innumerable uses and functions. To perceive is to discriminate between possible outcomes while attending to the environmental variables at hand. It has been argued that this attendance to afforded actions is an important perceptual process that is intended to reduce environmental uncertainty [Friston, 2022]. Overall, the affordance view establishes some important principles about the mind; it is necessarily situated, it is in a constant information exchange with an environment, and it is necessarily for action [Wilson, 2002].

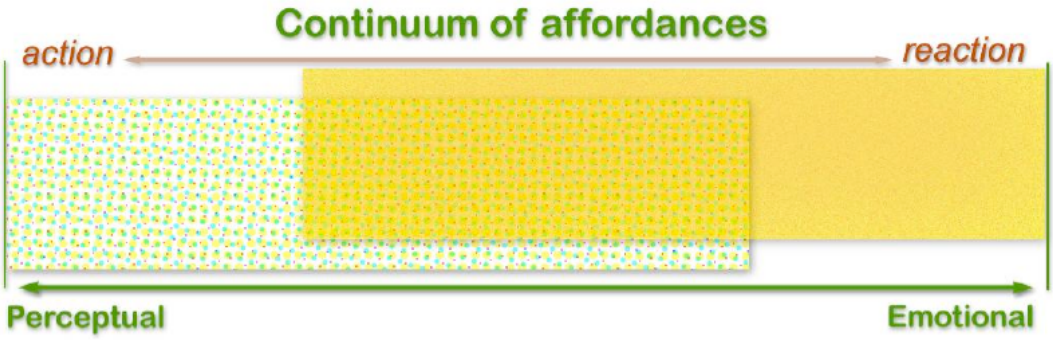


Figure 2.10: The continuum of affordances as described by Morie et al in the study of affordances in virtual environments [Morie et al., 2005]

The perception of an affordance must take place in tandem with a perceived event [Gibson, 1966, Gibson, 2000]. As outlined in the Gibsonian interpretation, this presupposes space and time perception to directly influence our capacity to perceive an affordance in an object. That is to say, the priors to affordance perception are a) our sensory organs b) spatial information and c) temporal information. However, other than these encodings, there are no other necessary priors for perception of affordances. Under this view, higher order cognitive processes are not necessary for affordance perception. The 'information pickup' is direct; the information is there to be perceived, and our sensual systems have evolved to perceive it. In sum, there is no necessity for representation, or for any other cognitive mediation, even if such abstractions can be shown to exist in and of themselves. They are not prerequisites for perception. This is known as the theory of direct perception [Djebbara, 2022].

Direct perception, and its related embodied accounts of mental phenomena, may be considered as an object-relative approach to perceiving objects; there are adapted mechanisms for perception and learned associations that enable us to perceive objects, relativistically, as relations between afforded actions. Emotion, in this framework, is about external themes and affordances, which are inseparable from the bodily position, action and condition that is tied to a given emotional quality [Hufendiek, 2017].

This approach can be understood as one that considers emotion, and the affordances that motivate emotions, to be intrinsically tied to agency and self, in that an emotion is a response to some meaningful content, that enables us to engage with the world in practical ways [Slaby, 2012]. In short, emotion, whether it is perceived in a face, or a voice, or the movement of a body, signal possibilities for action (i.e. affordances) in a way that contributes to our own selective responses to meaningful social perceptions [Kiverstein and Rietveld, 2012]. In this sense, perception, cognition, action and emotion may be considered as non-separate, fundamentally intertwined phenomena that disclose ways of skillfully engaging with the world [Jensen and Pedersen, 2016].

It is important to highlight, with regards to the discussion on realism that comes later, that Gibson's original conception of affordances was one of 'perceiver independent value', i.e. that affordances exist in the world independently of whether or a subjective agent is there to perceive them. However, in combining this notion with the framework of constructivism in emotion perception, the coherence in the present thesis stems from an interpretation of affordances as largely perceiver dependent, under the contention that perceivers play as much a role as an environment in the emergence of affordances, and as such empathy and direct perception are mediated by phenomenal experiences of affordances, in both spatio-temporal situations (e.g. climbing a rock wall) and affective social situations (e.g. observing an emotionally expressive virtual character) [Bogotá and Artese, 2022].

Before discussing empathy and direct perception, there is a final relevant point to be made about the affordances of these affective social situations. As discussed, in our perception of emotion, there are cues that may be used as affordances for the purpose of social affective behaviours. This claim can be placed under the 'Computers as Social Actors' paradigm [Nass et al., 1994] which suggests that our treatment of computers is fundamentally social, and is done so unconsciously [Nass and Moon, 2000]. The findings from our final experiment fall within the general scope of this paradigm, which suggest that users develop media-focused social treatments of virtual characters that differ from social actions towards real humans [Gambino et al., 2020]. This paradigm, when extended to include emotions, has been shown to influence perceptions of virtual characters emotions that differ between human and non-human controlled characters [de Melo et al., 2014]. The experimental work of this thesis would not dispute this, however there is evidence that non-conversational motor responses to emotions in our virtual character stimuli fall within the same definition as motor responses elicited by emotional humans, due to the affordances presented by the faces and voices of the characters. The Computers As Social Actors paradigm, by contrast, operates under a presumption of mental models for social conversational 'scripts' that are employed in

the linguistic treatment of computers. These scripts are proposed as test like functions, questions and modes of conversational inquiry, that are employed to explore the responsive nature of a machine. As described below, empathy as a perceptual phenomenon is not subject to social scripts. Instead, the treatment of virtual characters as social actors is proposed to be modulated by the direct perception of emotional affordances, presented via facial and vocal behaviour.

2.2.3 Empathy and Direct Perception

As well as a framework for understanding emotion, action and situated minds, direct perception can be applied to the understanding of social interactions. Specifically, direct perception comes into focus when applied to the concept of emotion perception. It is often posited as a divergence from more mainstream conceptions of emotion perception that categorise our ability to perceive the mental states of others as a form of simulation or mind-reading, that presupposes a mediating cognitive inference of a mental state, rather than a direct perception of it [Gallagher, 2008, Wiltshire et al., 2015].

A vital aspect of emotion perception is the notion of empathy, which has been defined in numerous ways. Our contemporary understanding of the word may be traced to the 1909 translation of Theodore Lipps term 'Einfühlung', which was used by him to refer to a 'region of knowledge' that involved the 'co-grasping' of 'knowledge about other selves' [Jardine and Szanto, 2017]. The word Einfühlung itself was previously used during the German Romantic movement by philosopher Johann Gottfried Von Herder as a metaphor for a method of inquiry, that included the component of 'imaginative reproduction of [another's] perceptual and affective sensations'[Matravers, 2017].

These early definitions are largely similar to modern scientific descriptions of empathy. However, the debate on empathy requires a broader scope than that provided by any definition. Like emotion perception or perception in general, empathy is an active process, as much as it is also a quality which defines the character of social experiences [van Dijke et al., 2020]. This combination of action and quality may be conceptualized in a variety of ways, with the domains of cognitive neuroscience and psychology often invoking the perception-action model [Preston and de Waal, 2002, Preston, 2007] in order to explain the mechanics of this phenomenon. The proposition goes that the perception of an emotion and resulting empathic response involve a shared representation between the emotional agent and the perceiver. By 'representation' Preston et al. [Preston, 2007] mean a pattern of behaviour and neural activation which is reliably repeated under the same emotional conditions. This shared representation is contended to be translated into a form of action, be it the activation of other neural circuits, the

shedding of a tear or the reaching out of a hand for a comforting touch. This representation can incorporate direct perception models of empathy, as well as mental imagery or simulation [Preston, 2007]. In contrast, the direct perception model contends that no such representation mediates our access to the minds of others, which is instead acquired through a social understanding that necessarily takes place experientially rather than through any kind of mental projection [Zahavi, 2011, Gallagher, 2008].

In light of our partial explanations of empathic processes, questions must be asked about its relationship to emotion perception, and the extent to which digital simulacra of humanity can influence empathic responses. Answers to some of these questions are proposed as the main contributions of this thesis. However, one important takeway from theoretical approaches to empathy and direct perception is that responses to emotion are contingent upon activity, as much as, if not more than, mental models. In other words, the activity of another agent, virtual or otherwise, modulates my attention processes in a direct way; I perceive the mental state of another by virtue of my actions and their actions; I scan their face, I listen to their voice, and in this way I empathize, mentalize, and in a very direct manner I construct the emotional meaning of their features, and produce an empathic response [Kiverstein, 2015].

In differentiating between models of empathy, the studies described by this thesis do not support any explicit account of empathy, such as Simulation Theory, as defined by the imaginative 'simulation' of another's mental state, [Goldman, 2002], but is suggested to support the specific condition of vicarious experience of another's mental state [De Vignemont and Singer, 2006], which can fit with both direct and narrative models of empathy, as well as simulationist models [Gallagher, 2012]. The subtle distinction between these two accounts being 'imaginative simulation' and 'vicarious experience' may seem arbitrary, but can be considered important when we take the consequences of either approach; namely that for simulationism, the experience of empathy is a thought-out cognitive process, whereas the direct perception model considers it to be perceptual, a felt sensation rather than a cognitive representation. There is a direct relationship between a perceiver and an emotionally expressive agent that is manifested in actions that are dependent on direct perception of affordances, rather than an imaginative reproduction of another mind. This proposition is based on our findings which suggest motor empathy for virtual characters, with the process of cognitive empathy (i.e. the attribution of mental states to others) heavily dependent on the inclusion of a real human voice. We believe that our findings on facial mimicry in particular, namely the event where there is emotion-specific muscle activity in the faces of both virtual characters and users who observe their activity, highlight direct experience of the emotional states of our virtual characters, that may be understood as a basic mechanism for empathy that

can enable shared affect and joint attention to take place [Kiverstein, 2015]. With this theoretical approach, we may treat motor mimicry as a perceptual mechanism rather than a representation.

Given that this thesis was focused on how afforded emotions are perceived by users, the gaps in the literature on emotions and affordances of virtual characters are considered to be at the core of the main research questions. Specifically, the niche that was to be explored in our experimental work is how the perception of affordances and emotions, as well as corresponding empathic responses, could be dependent on the realism of a virtual character. This gap in the literature is treated as directly stemming from both the capacity of a user to construct instances of emotion and the affordances of the virtual character that they perceive. As such, the literature described in the sections above are necessary theoretical components to accurately describe the observed effects.

In conclusion, there is also a need to clarify the relationship between direct perception models of empathy and the aforementioned theory of constructed emotion. It is an assumption of this work that requires some clarity, although the link itself is intuitive, as implied by the previous paragraph; the manner in which we construct the instance of an emotion in the behaviour of another agent relies on a direct perception. The pretense of an internal representation of emotion is by comparison more akin to the basic emotions theory, whereby there are idealised versions of emotional expressions that we as agents possess, and these representations may be matched to instances of emotion that occur in our day-to-day lives. Yet, while we can imagine such a representation, and attempt to match it to a given emotional stimulus, that does not appear to be what we are doing when we actually perceive an emotion in the wild. Our perception of emotions and our corresponding empathic reactions, whether they be for a real suffering human being or a happy virtual character, are in fact much more active. We are directly involved in the matter of understanding the emotion, which requires more than a representation and consequent action model. It requires us to act out the emotional expression we perceive, while engaging with its expressive agent. In the experimental work described by this thesis, these perceptions are defined by the content of a virtual characters face and voice.

2.2.4 Face and Voice Processing

In order to understand how emotional expressions in virtual characters may afford both empathic responses and the perception of emotion in those characters as a whole, we must first cover how human beings perceive emotional faces and voices. There are numerous ways to perceive emotion in social interactions; the internal state of an emotional

agent, human or non-human, is largely invisible to a perceiving agent. Their internal state can be communicated via emotional content through a diverse range of channels (See Figure 2.11), which can include both facial and vocal information, as well as bodily behaviours and other channels of communication such as physiological responses or text-based symbols [Lange et al., 2022].

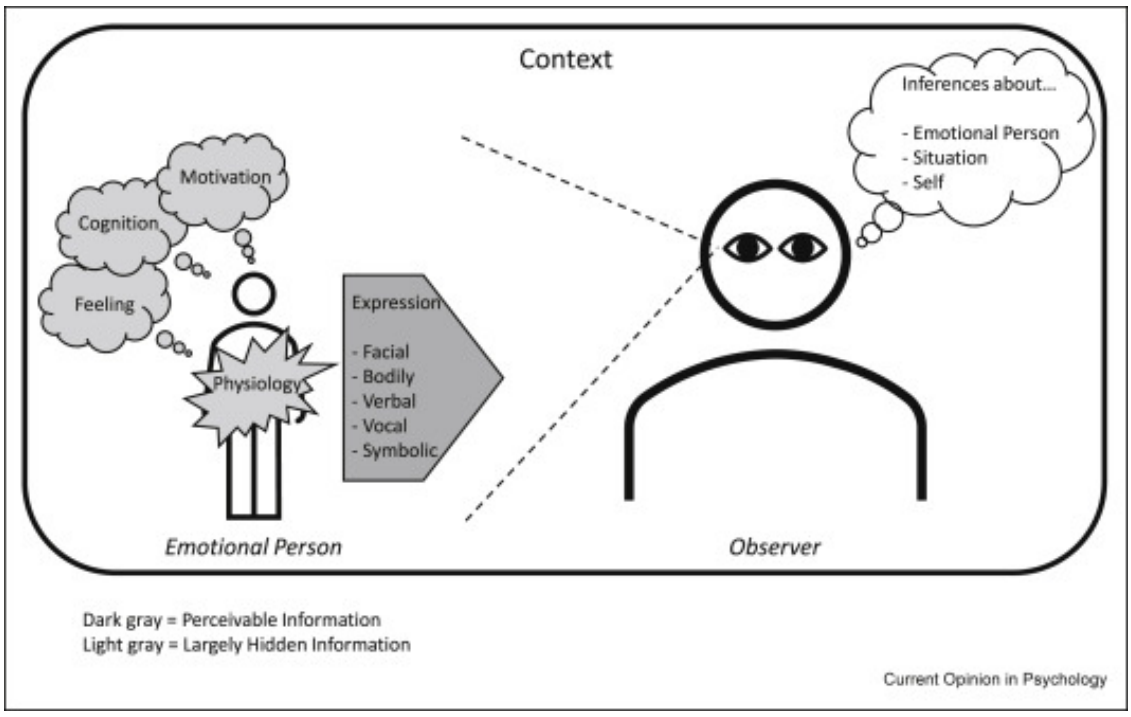


Figure 2.11: A useful schematic for describing the different ways an emotion may be expressed and perceived [Lange et al., 2022]

.

While it has been established that, in the scanning of a visual scene, faces tend to be given the foremost allocation of attention compared to other objects [Hershler and Hochstein, 2006], we also tend to experience illusory perceptions of faces in situations where they are not present, by creating a recognisable facial pattern from a composition of non-facial patterns and stimuli [Zhou and Meng, 2020]. It has been argued that this latter tendency reflects a perceptual mechanism, rather than a cognitive-representational one, which has been adapted for the purpose of prioritising social information from faces, specifically that of social attention [Palmer and Clifford, 2020]. Faces are generally considered to be vital for communication purposes, where expressions can be deployed to emphasise a point, regulate conversation dynamics, and convey particular emotional states [Pelachaud and Bilvi, 2003].

Neuroscientific approaches have highlighted the visual affordances of faces to be processed via an occipital-to -posterior parietal brain area stream that has been associ-

ated with visuomotor rather than just visual perception [Debruille et al., 2012]. In other words, faces are processed as objects for action rather than as objects that play a passive role in a perceptual schema. While this may seem intuitive, it is important to note that numerous other brain areas have been associated with face perception tasks, and the distribution of this processing over many other regions and neuronal networks may attest to the multifaceted range of actions that faces can afford [Zebrowitz et al., 2011]. One of these affordances is the emotional expression of the face, which can afford further social action mechanisms to take place [Zebrowitz and Montepare, 2008] and which may be processed in tandem with other cues to build the impression of an emotion, and hence, perceive it. This perception has been observed to take place rapidly, prior to conscious evaluation of the affective content of the face [PALERMO and RHODES, 2007]. Indeed, these processes are proposed to be mediated by the social affordances provided by faces, observable in the neural responses to both real and fake faces [Debruille et al., 2012] (See Figure 2.12).

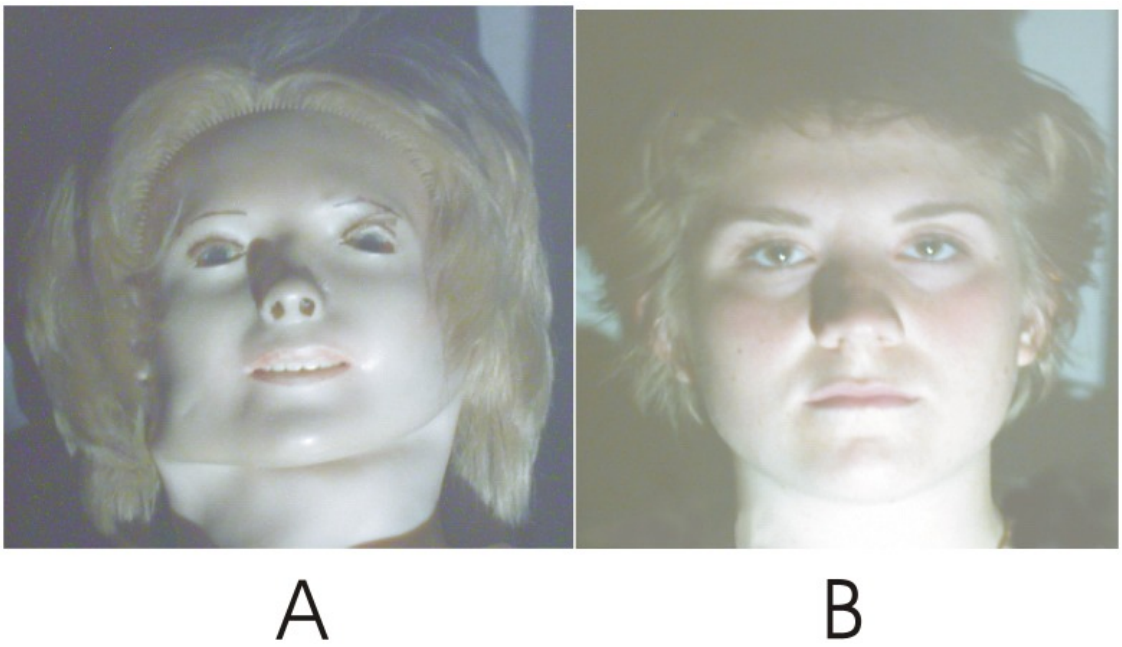


Figure 2.12: Dummy and real faces used to study the neural responses to the social affordances provided by faces [Debruille et al., 2012]

Additionally, for most of our experiments, the emotional faces were almost always accompanied by emotional voices. Comparatively less work has been done on the topic of voice processing, although the same neuroscientific methods expounded the role of the temporal and auditory cortices in the perception of a voice [Schirmer and Adolphs, 2017], with prosodic features of voices that are specific to emotional voices observed in experimental settings observed to be inhibited by visual stimuli [Mothes-Lasch et al., 2011]. Indeed, such are the range of characteristics that are perceptible in the content of both faces and

voices, such as emotion, gender, age etc, that the latter can be regarded as an 'auditory face' [Belin et al., 2011]. This metaphor may be extended to include the notion that voices are given attentive and neural salience ahead of other sounds, the way faces are proposed to do compared to other images [Ethofer et al., 2009]. In relation to emotion perception, it is important to note that vocal expressions are necessary to ground the ambiguity of facial expressions; facial expressions alone are context dependent (as noted earlier, a smile may be threatening) and so the use of vocal utterances are vital to outline specificity in the communication of emotions, and thus reduce uncertainty as to what emotion should be constructed by a perceiver [Lindquist and Gendron, 2013].

Nonetheless, strong evidence exists to suggest that, while voice and face information are often combined to augment the processing of content in either, faces tend to receive more attention when it is incongruous with a given voice, i.e. when their content does not match [Stevenage and Neil, 2014]. While, as will be discussed in the next section, this incongruity has been observed to influence reactions to human-like virtual characters, evidence also exists to suggest that, for the perception of emotion, facial information is sufficient to transmit both valence and arousal of emotion in virtual characters, while the voice becomes more salient once non-expressive faces are used [Torre et al., 2024]. Importantly, this emotional information has been recorded to have a greater effect on users who perceive a virtual character as a human controlled avatar rather than an independent agent [de Melo et al., 2013]. While bodily information and posture have also been observed to influence emotion perception and are an important element of emotion expression [Lange et al., 2022], there is a consideration that faces and voices together are sufficient for emotion communication, with voices in particular necessary due to their use in explicitly defining what emotion is occurring for a perceiver to construct the correct sense of a facial expression [Lindquist and Gendron, 2013].

The most important message of the above cited works is that the perception of faces and voices normally occurs in tandem [Stevenage and Neil, 2014], rather than in isolation. We do not compare facial and vocal information to a general mental representation of an emotion that is activating in some innate cognitive function. We see a face, its eyes, its mouth, the creases in its forehead, and we use those signs to make judgements as to the emotional qualities that are afforded, and how they might affect our next judgements, actions and predictions. Similarly, when we hear a voice, we hear its volume, its prosody, its pitch and intonation features, and through those, normally combined with facial information, an emotional expression is perceived directly. This conception of afforded cues in faces and voices is addressed in the literature on virtual characters, but rarely with particular reference to the construction of emotion. In formulating these aspects of virtual characters as an element of human emotion perception, this work is proposed to

unify scientific perspectives on both topics.

As this thesis focuses on how realism modulates the perception of emotion in virtual characters, the following section is dedicated to defining the terms of realism perception that motivated the development of our experimental works. Ultimately, these theoretical foundations lead us to our conclusions on how the combination of visual and vocal realism can afford empathy for virtual characters.

2.3 Perception of Realism

This section is primarily devoted to describing our perception of realism, and attempts to situate the notion of realism within the contemporary scientific literature on perception.

Firstly, the most important term used in this thesis is that of perception. Perception has traditionally been understood as the process by which objects are revealed to each of us through the operations of our senses, which are supposed to allow us to interact with the real world [Merleau-Ponty, 2004]. However there is room for debate in this definition, given that there can be no coherent descriptions of our perception of 'virtual' objects (i.e. objects that do not exist in the real world) without some concessions as to the reality of the experience. In other words, we perceive and react to virtual objects in the same way as real ones, rather than fictionalising those same perceptions or categorising them as separate ontological structures compared to perceptions of real objects [Chalmers, 2017].

Realism is a broad term, and so in order to be understood properly it must first be approached from a broad perspective. For example, one might say that an orange, placed on the table in front of you, is real. It exists, it has a form of its own that is perceiver-independent, and hence, real. In this case, the term 'perceiver-independence' denotes a separate existence, whereby the orange is there, existing in independent reality, whether or not you are there to perceive it. In this case, the degree of realism is apparent in the colour, taste, and weight of the orange when you see it, eat it or hold it. Its realism is intuitive, even obvious. It is in the world. This is arguably how we perceive real people, as independently existing agents.

Similarly, realism may be considered an aspect of a photograph or a vivid lifelike depiction of an orange. If the photographer or painter depicts the orange carefully, with utmost accuracy and attention to detail, one might say that there is realism inherent to these depictions of oranges. The realism is there, in the style of the creation. There is an action from the creator, who deliberately affords realism in the crafted object. And then there is action from the perceiver, who observes the piece and comes to assess it in some capacity or another, as realistic. In these cases, the realism cannot be perceiver-independent. Photographs of oranges, and realistic paintings of oranges, are for perceiving. They depict nothing, unless there is someone on the other side of the canvas perceiving the realism of the object. Realism in this case is perceiver-dependent. This is the manner in which we perceive virtual replications of people; they too are designed for the purpose of perception.

However, what if we were to claim that the orange (the real one) on the table in front of you was, in fact, also perceiver-dependent. That is to say, if you were not there to see it,

to taste it, or to hear it bang against the table as you hammer it in existential frustration, it would not be there, and it would not be 'real'. Given how each and every person who reads this, and most who don't read this, interacts with the world in a way that relies on objects (oranges, hospitals, computers, factories, phones, other agents etc.) to be there when we are not there to perceive them, it may sound like utter sophistry to make an argument of this kind. We know things are there when we aren't looking at them. Human beings have an 'innate' capacity to know this. And, unless another agent or a disorder is involved, this faculty is altogether very reliable.

Two theoretical advancements in twenty-first century cognitive science give us good reason to doubt this intuition. The first paints a general picture of our mechanistic interactions with the physical world that is very different from the sense we have of it. The second consolidates this counter-intuitive claim in the field of emotion research, and is especially pertinent to the work described in this thesis.

The first advancement comes from a peculiar combination of specialty domains, namely those of quantum physics and evolutionary game theory, known as the interface theory of perception [Hoffman et al., 2015, Hoffman, 2024]. The advancement in question can be expressed in the form of some simple premises. If we cannot directly perceive, for example, quarks (a type of quantum particle), then we cannot perceive the underlying, fundamental structures of reality. This is self evidently true, otherwise we would not need a large hadron collider to study quarks. Since this is true, not only are the basic constituents of our phenomenal experiences (for example, the arrow of time) mere artefacts of this underlying structure, but we cannot even access this underlying structure. Hence, not even the arrow of time is a real, perceiver-independent phenomenon [Hoffman, 2024]. What then, are we perceiving? Well, evolutionary game theory suggests that we perceive and interact with icons, mere interfaces, that give us enough useful information to take advantage of the evolutionary algorithm. In a sense, our perceptions are like clicking around on a desktop. We see only the functions of the computer, not the logic gates [Hoffman et al., 2015].

The point that is made here is the blurred distinction between real and unreal that is highlighted by our general interactions with objects in the world. What the work above suggests is: *All things that we perceive are some kind of affordance for acting in the world*. In our day-to-day perception, we perceive what is useful, not what is real.

This is important to the second theoretical advancement of note for this study. If none of the basic physical constituents of our existence are perceiver-independent, then what about the other constituents? For instance, is there such a thing as a virtual character? Or is that merely a label, born of misunderstanding the consituent parts of a given

stimulus, say, the complex interplay between a matrix of polygons, a motion tracked animation and a recorded human voice? This thesis attempts to address these questions by a) specifying specific and measurable forms of realism (i.e. affordances that we perceive in virtual character contexts that are dependent upon the perceiver) and b) identifying which of these affordances can be viewed as primarily necessary for providing emotional information through these virtual characters. To this end, this notion of realism is proposed to allow us to view specific forms of realism, such as vocal and visual realism, or movement realism, behavioural realism, perceptual realism and so forth, as perceiver dependent values that are expressed by objects, which allow us to perceive them as real. That is to say, for example, when a virtual character moves, we perceive its movement, and we are constrained by the subjective nature of our perception to classify it in some sense, be it moving as a human moves, or moving in an erratic or unrealistic sense. The work of this thesis addresses realism at this intersection between perceiving-as-real and perceiving-as-unreal.

A related line of questioning might be: is there such a thing as an emotion? Do we experience emotions? Do we perceive them as they are? Or do we construct the facets of an emotion with our minds? Do we create them, as a painter creates a realistic orange? Or do we use them, as a painter uses a brush to paint her orange? With these questions in mind, the task of outlining which aspects of virtual characters, oranges or anything else are inherently real or constructed, appears to be a more significant task than at first it might have seemed.

However, there is room in these issues to narrow the focus of this thesis into a more definite shape. The second aforementioned theoretical advancement in relation to 'constructive perception' comes from the field of emotion research, which has in recent years been pivoted towards constructivism [Barrett, 2011], in place of the prototypical 'basic' emotion theory [EKMAN, 1999]. Of course, as with all philosophical dichotomies, there is reason to believe that both theories contribute huge amounts of valuable knowledge to the field of emotion research, and highlight some real and important findings that advance the field [Karnaze, 2013]. While the basic emotions will be highlighted in great detail in this thesis, it is important to first highlight the contribution of constructivism to form the theoretical basis of this work.

As with the interface theory of perception (discussed above) from [Hoffman et al., 2015], constructivism may be considered a post-Darwinian account of information processing, in this case specific to emotion, that deals with variation within populations as a perceptual challenge that humans have evolved to overcome. In this light, supposed instances of emotion expressions are perhaps challenged themselves by this principle of

variation; if there are differences within populations in terms of expressing and perceiving emotions, then what is it about emotions that is objectively real? The answer may be just that it is the experience of the emotion; signalled by a flush, a clenched fist or a shiver. It is not in the expression of the emotion by another person that we come to know the reality of emotion. We construct emotions in the behaviour of others, in an appraisal that is based off personal experience and the variation on that experience that we learn from others to construct into a similar shape [Barrett, 2013, Barrett, 2016, Barrett, 2011, Barrett et al., 2019]. This is the realism of emotion. Perception involves the construction of this reality, and whether we understand a smile to be threatening or joyful is largely dependent on this capacity for construction.

Yet, to focus on emotion in virtual characters we must also deal with the notion of realism. As outlined above, the term realism is problematic, given that things we regularly treat as real, such as emotions or oranges, may in fact involve a good deal of perceptual confabulation in order for them to be perceived. In order to measure whether the same is true of our perception of virtual characters, the aim of the the following three subsections is to reduce this notion of realism to specific properties that could be manipulated in experimental formats. By including these contemporary views of realism in both emotion perception and perception in general, this thesis aims to augment existing theories on realism perception using the foremost theoretical advances in both fields. In doing so, this work is aimed at advancing the understanding of virtual characters from a broader perspective. The interface theory of perception and the constructivist view of emotion both directly relate to the core research questions; when we perceive realism and emotion in virtual characters, we are employing social mechanisms to construct emotions using afforded qualities that are useful if they are treated as 'real'. The consolidation of these theories with the study of virtual characters is one additional gap in the literature that was targeted by this work.

The complicated story of our interactions with the world must be reduced in order to be measured. In studies of virtual characters, the term realism can be used to refer to many different properties; perceptual realism, social realism, visual realism, vocal realism, interaction realism, behavioural realism and movement realism all serve as prominent examples which have been analysed in the literature [Wages et al., , Zell et al., 2015a, Zibrek et al., 2021, Zibrek et al., 2018]. Each of these aspects, when considered under the auspices of the interface theory of perception and the constructivist theory of emotion, represents an element of virtual characters that is developed to afford something that is conceptually familiar, something that appeals to our minds in a manner which mimics the salient loci of our perceptions. In other words, these elements of realism contain properties which lend themselves to the

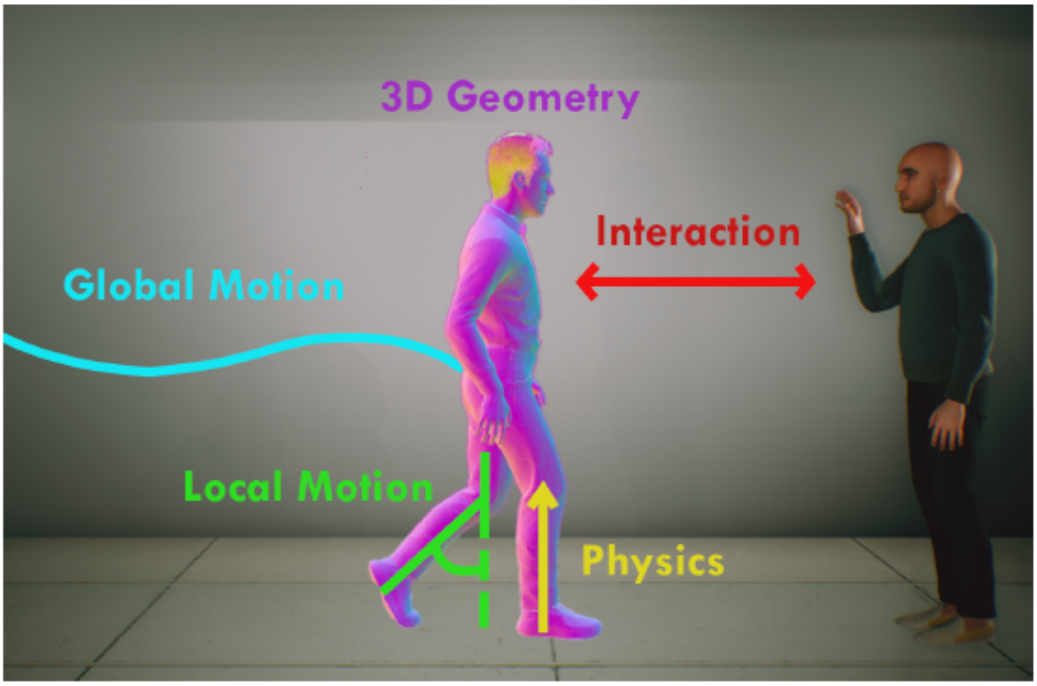


Figure 2.13: Types of realism in the animation of virtual characters as defined by Rekik et al [Rekik et al., 2024]

enactive, constructive faculties of our minds that are described by the theories explicated in this section.

The focus of this thesis was on the application of advanced computer science techniques to study reactions to realism in virtual characters, particularly with reference to their human-likeness; how do we perceive emotions in virtual humans that more or less resemble real humans? The subections below highlight the twin factors of visual and vocal realism as they relate to virtual agents, and how these forms of realism may be measured to impact our emotional responses. Our visual realism definitions comprise of a combination of geometric and illumination realism (as described below) while vocal realism is considered to be a form of interaction realism (See Figure 2.13) [Rekik et al., 2024].

2.3.1 Visual Realism

Visual realism can be understood as a key issue in the field of computer graphics, and may be broken down into two components; geometric realism and illumination realism [Slater et al., 2009]. Geometric realism is defined by the number of polygons, which are the simple base shapes that make up the model of any 3D model of a virtual character. For many modern character models, this geometric complexity can be achieved using a greater number of polygons, which allow for more complex detail in the plane that the polygons occupy [McDonnell and Mutlu, 2021]. Illumination realism is a different but

related method of generating visual realism in virtual characters, with modern technological advancements allowing for many developers to employ ray tracing for real time animations, as well as texture mapping techniques for adding colour information to the geometry of each pixel used to make up a virtual face. Contemporary ray tracing methods focus on effectively simulating the movement of light across a virtual character, with the simulated light paths designed to computationally mimic real-world light based phenomena such as reflection and shadows, the details of which advance the visual realism of a virtual face as it is animated and in motion [McDonnell and Mutlu, 2021].

In general, the study of visual realism in virtual characters has provided numerous technical and theoretical advances in the fields of computer science, psychology and human-computer interaction. While there is evidence that computer generated faces are subject to similar perceptual and neural processing mechanisms as real faces [Moser et al., 2007], research also suggest that semi-realistic faces cause a sense of unease in human experimental participants [Kätsyri et al., 2017a].

The application of visual realism to virtual characters is an important consideration for designing human-computer dialogues [Jung et al., 2011], and has been argued to improve communication in pedagogical settings [Gulz and Haake, 2006]. Nonetheless, the value of visual realism in virtual characters is hotly debated [Beck, 2008, Wang and Doube, 2011]. This debate centers on justifying the computational cost of highly realistic characters, which may ultimately hinder user engagement rather than augment it [Kunz, 2015].

Justification aside, there exists some good evidence that visual realism in virtual characters influences brain activity in a way that is unique to realistic human likenesses [Schindler et al., 2017, Mustafa et al., 2017]. On this basis, there is sufficient reason to question whether realistic human likenesses may influence other behaviors at the physiological level. While the experimental work of this thesis implies that there is a degree to which we treat virtual characters as we do real humans, this unique response to realistic human likenesses may be viewed as a sign of different afforded actions. The uniqueness of EEG amplitudes for realistic virtual faces compared to real faces as defined by [Mustafa et al., 2017], as well as the differential effects on EEG amplitude across stylized, realistic and real human faces expressing positive and negative emotions (See Figure 2.14) recorded by [Schindler et al., 2017] can be considered physiological observations that are relevant to the aforementioned theories of interface perception and constructivism. In particular, the N400 component measured by Mustafa et al is commonly associated with mismatched expectations, and had a significantly higher amplitude for realistic computer generated characters compared to less uncanny real humans. In per-

ception terms, this suggests that the actions afforded by a face, which can be considered an icon for social action, are predicted to be those of a human agent before that prediction is corrected, through the identification of cues which define the character as computer generated. Similarly, the fact that real negatively valenced characters (angry faces) evoked the greatest activation of the N170 component as measured by Schindler et al implies that the construction of the concept of 'anger' is more salient when viewing angry humans compared to angry computer generated characters.

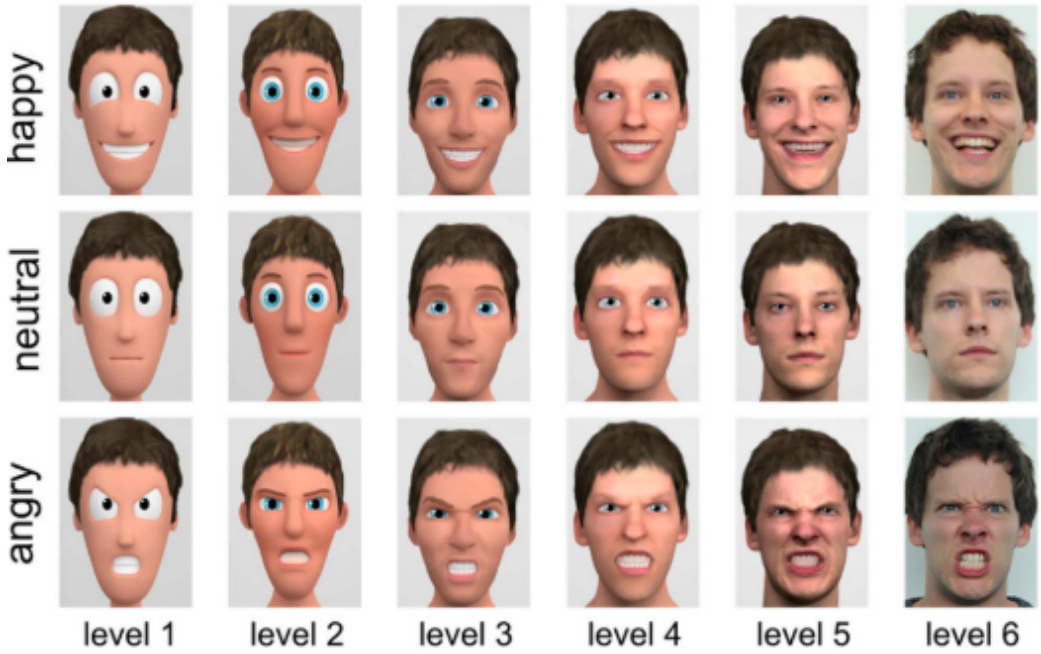


Figure 2.14: Example of differing levels of visual realism, used by Schindler et al [Schindler et al., 2017] to study realism and emotion perception.

Previous research in these areas has shown visual realism in animated virtual humans (aka photorealism or appearance realism) to be influential over emotional responses to realistic agents, with realism provoking greater excitement in virtual reality [Zibrek et al., 2019], as well as more distress for a virtual patient in screen based studies [Volonte et al., 2019]. Moreover, perceptual studies in this area have demonstrated that realistic characters have been considered more trustworthy [Seymour et al., 2019] and more comforting [Araujo et al., 2021] compared to unrealistic characters, in both VR and screen based contexts.

Visual Realism for Avatar Communication

Studies using realistic virtual avatars to represent individuals in both immersive and screen-mediated virtual spaces have benefited from advancements in face tracking technology. The use of real-time facial mapping has been shown to induce SoE with avatars

under the control of individuals [Kokkinara and McDonnell, 2015]. This effect can be directly related to the realistic detail of avatar faces, given that a reduction in facial detail and real-time responsiveness has an antithetical effect on subjective reports of these psychological dimensions of avatar control [Kokkinara and McDonnell, 2015]. The ownership dimension of SoE in particular has been demonstrated to benefit from improved realism in avatar interactions [Gorisse et al., 2019].

There is some evidence to suggest that the use of avatars for collaboration purposes can be improved with visual realism in Immersive VR. Some studies have implicated full-body avatars as the most useful for generating senses of social presence between users in Social VR settings [Wang et al., 2019, Heidicker et al., 2017] and highlighted their impact in enhancing engagement of users while doing a virtual task [Fribourg et al., 2018]. However, similar results have been achieved with head-and-shoulder style avatars [Yoon et al., 2019] provided they can represent a users bodily movements to a sufficient degree. These effects can even be observed without representing an individual as their avatar in realistically detailed appearance [Gonzalez-Franco et al., 2020].

2.3.2 Vocal Realism

In addition to visual realism, our work aimed to evaluate the effect of voice realism and its affordances for emotion perception. The synthesis of realistic voices has been subject to rapid technological advancement. In recent years it has become possible to generate voices that can be modified to include specific features, such as identifiers of age, gender and emotional state. While these sound manipulations can create artefacts which may be perceived as unnatural, there are now both audio and audiovisual speech synthesis systems that are designed to produce more realistic 'phonemes', which are the sounds that are used to make up the particular noise that distinguishes a given word [Mattheyses and Verhelst, 2015].

Higher quality speech synthesis systems have been developed to generate speech using both signal modelling and articulatory modelling approaches, which are focused on modelling the signal of a sound and the method of its production respectively. Additionally, contemporary neural networks can be trained on datasets of human speech, in order to generate convincing replicas from text [Raghavendra et al., 2010, Zen et al., 2013, Bendel, 2019]. Such techniques are often aimed at producing more human like speech through the use of vocal prosody. Prosody is a 'paralinguistic' aspect of human speech, in the sense that is not about the actual words used in a vocal utterance, so much as 'the way they are said'. In other words, prosody can refer to specific computable elements

of an utterance, such as pitch, timing and volume [Crumpton and Bethel, 2016]. It is also important to note that prosody is considered to be a key aspect of vocal emotion expressions, which can be specifically expressed through pitch, timing and volume, or through the use of other qualities such as intonation, emphasis and tune [Taylor, 2009]. For example, a small pitch range (the difference between highest and lowest pitch during a vocal utterance) is often associated with sadness, while loudness of a voice can be indicative of happiness or anger [Crumpton and Bethel, 2016].

In terms of 'realistic' synthesised voices, there have been repeated efforts to develop computational models that are capable of analysing large datasets and using them to develop features that are necessary for creating a human-like voice, such as hidden Markov models that can generate specified parameters of pitch and spectrum [Yoshimura et al., 1999]. In addition, more recent models have been geared towards transforming relevant data in a given speech dataset, in order to generate more expressive voices using vocoder principles, adaptive prosody models and copy synthesis techniques for imitating vocal tract activity [Turk and Schroder, 2010, Kaneko and Kameoka, 2018, Shi et al., 2023].

However, despite these recent advancements, the synthetic voices of speech interfaces are generally treated as more formal and less authentic than real human voices [Doyle et al., 2019]. When applied to virtual characters (as in Figure 2.15), numerous recent studies have focused on the application of synthetic voices. For example, voices in virtual characters have been measured to benefit from self-similarity (i.e., the voice is less synthetic and more like the user's voice) when it comes to believability, but are also considered significantly more eerie [Kühne et al., 2020, Guo et al., 2024]. Although human-like synthetic speech can have this effect of eeriness, it has been shown that synthetic voices are considered less likeable compared to real voices [Baird et al., 2018, Ferstl et al., 2021b] as well as being considered easier to understand and generally more expressive [Cabral et al., 2017] and of generally lower quality than real voice [Parmar et al., 2022] . There are also shortcomings with regards to including synthetic voices in public radio broadcasts [Scott et al., 2020] as well as in audio-books [Rodero and Lucas, 2023], where human voices are also preferred.

While synthesised voices have been explored in relation to the effect they have on human perception, many aspects of social behaviors directed at highly realistic virtual agents with unnatural voices have not been fully explicated. It has been shown that a mismatch in the realism of human face and voice can create a feeling of unease and discomfort in experiences of quasi-realistic virtual characters [Mitchell et al., 2011b]. It is on this point, where the combination of vocal realism together with visual realism influences our emotional responses, that we begin to close in on the core research themes of this thesis,



Figure 2.15: Virtual characters in human and robot form tested in conjunction with real voices and text-to-speech generated voices [Ferstl et al., 2021a]

given that both human faces and voices are important for the expression of emotions [Potard et al., 2016]. The mismatch in afforded realism between these two modes is proposed to produce an uncanny effect, as outlined in detail during the next section.

Some variations in vocal realism, between real and text-to-speech voices in isolation, have also been observed to produce an uncanny effect, but at the same time have challenged the previous studies which have found a preference for real as opposed to synthetic voices [Abdulrahman and Richards, 2022]. This was some motivation for exploring synthetic voices with realistic virtual characters. However, work on assessing emotional responses to synthetic vocal agents has been beneficial to the study of synthetic voices, with some Bayesian approaches producing evidence for a 'habitability gap' that highlights the necessity for aligning vocal and visual affordances in the production of intelligent virtual agents [Moore, 2012, Moore, 2016]. This finding emphasises the importance of matching the levels of realism in voice and appearance, which is similarly relevant to the study of the Uncanny Valley.

2.3.3 The Uncanny Valley

The meaning of the word 'Uncanny' is often associated with the work of Sigmund Freud and his essay entitled 'Das Unheimliche", which translates roughly to the phrase 'unhomelike'. In this essay, Freud takes a psychoanalytic approach to the field of aesthetics, whereby the phrase unhomelike (later translated into the word uncanny) is understood to refer to the disquieting recurrence of the familiar [Freud, 2017]. This same essay also invokes the short story of German artist ETA Hoffman and his story 'The Sand-Man', where one of the main characters is described with elements of inanimacy, mechanistic stiffness and deadness in the eyes. The core definitions in this work ascribe a kind of dialectical opposition between familiarity and unfamiliarity that can be seen in the use of the words canny and uncanny; in essence the term uncanny can be used to refer to

something that is familiar and yet unfamiliar, eerie yet appealing, or a combination of the natural with the unnatural. This kind of terminology is often used today in measuring the psychological phenomenon known as the Uncanny Valley.

The theoretical notion of the Uncanny Valley [Mori, 1970, Mori et al., 2012] has pervaded scientific discussions of humanoid robots and virtual humans since the end of the 20th century. The term describes a state of unease induced by an artificial representation of a human that occurs at a level of realistic portrayal where that representation differs only subtly from a real human.

This proposed valley begins at the point where people typically begin to report negative feelings towards a virtual or robotic human. The valley is considered to end at the point where realism and positive impressions resume their correlative increase (See Figure 2.16). This proposed correlation implies that there is an upper boundary of realism where both likeability and realism are at an optimized level, although Mori originally supported the idea of designing to stop short of the valley rather than traversing it through maximal realism [Mori, 1970, Draude, 2011].

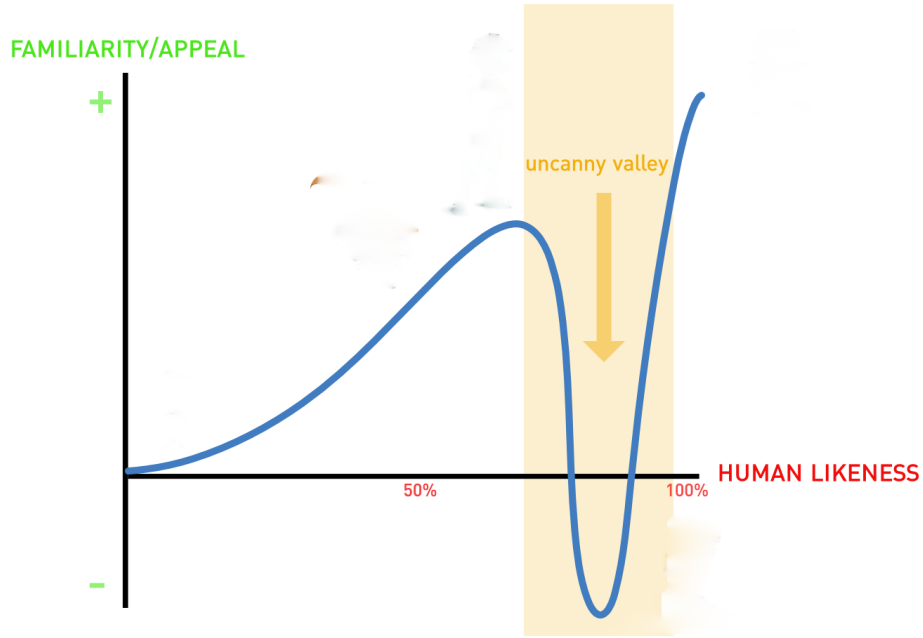


Figure 2.16: The proposed dimensions of the Uncanny Valley

While sometimes considered to be a theory lacking in substantiation [Blow et al., 2006], there has been growing evidence for the Uncanny Valley effect fostered by studies which have implicated the realism of characters as a governing influence on impressions of virtual characters [ichiro Seyama and Nagayama, 2007].

With the proliferation of high quality animation technology, however, the correlative

desynchronization of likeability and appearance realism has proven to be extremely relevant for researchers seeking to contribute to animated characters in films, cartoons and games [McDonnell, 2012], where audience reactions bare importance for the design of virtual humans.

For instance, stylized cartoon renderings of virtual characters have been rated more favourably than more realistic but sickly looking counterparts [Zibrek and McDonnell, 2014]. Realistic virtual humans that exhibit low behavioural realism appear to provoke similar effects, compared to high-realism conditions [Guadagno et al., 2007]. There is also evidence that more realistic animations of humans may evoke less trust than cartoon versions [McDonnell and Breidt, 2010], and are more easily categorised by humans than semi-realistic characters [Kätsyri et al., 2017b].

Moreover, investigations have highlighted the roles of character appearance and behaviour [Bartneck et al., 2007], individual subjective traits [MacDorman and Entezari, 2015] and facial expressions [Tinwell et al., 2011] in shaping the experience of virtual humans, in efforts to quantify a location of the uncanny valley on the scale of possible human representation. However, the work of [Bartneck et al., 2007] suggested that the Uncanny Valley may not have an endpoint, and may in fact represent a preference for non-humanlike characters, although their study was conducted with static images as opposed to dynamic ones. Another finding, of of great import to the focus on subjective traits in this thesis, come from the above cited study by [MacDorman and Entezari, 2015], which found that the Big Five Personality trait of Neuroticism and the Interpersonal Reactivity Index trait of Personal Distress were both measured to predict the perception of the Uncanny Valley.

Many such studies involve full-body representations of virtual humans, although weaker uncanny valley effects have also been measured with screen-mediated images of computer generated faces [Kätsyri et al., 2019].Numerous studies have also sought to model human behaviour to methodically describe the effects denoted by the uncanny valley and their perceptual basis [Schindler et al., 2017]. Some proponents of the theory point to evolutionary preferences for aesthetics [Hanson et al., 2005] as a culprit in the generation of negative character impressions. An evolutionary perspective can be supported with evidence that the uncanny valley has been observed in the visual behaviour of monkeys; when presented with stimuli representing real, synthetic and near-real synthetic monkey faces, they spent significantly more time looking at the synthetic and real compared to the realistic synthetic faces [Steckenfinger and Ghazanfar, 2009]. Considering uncanny valley effects from an evolutionary perspective can be useful for developing guiding prin-

ciples in the domain, but does not provide a framework for measuring perceptual effects of vocal or visual realism specifically.

Another approach proposes concept categorisation as the prime culprit for perceived uncanniness [Tondu and Bardou, 2011]. The effects denoted by the uncanny valley have been proposed to arise from category errors [Feldman et al., 2009] where difficulties discerning the difference between a human and a replica version may lead to a state of unease. This notion of category error or category uncertainty has been corroborated by studies which have compared category sorting tasks to negative impressions of animated faces, where slower sorting time for morphed faces were correlated with lower likeability ratings [Yamada et al., 2013]. However, this approach was not considered appropriate for framing effects of character emotion on studying the uncanny valley.

A more applicable framework can be derived from studies which consider inconsistent levels of realism between features of a virtual human as the factor that may explain why negative impressions begin to arise at high accuracy representation [Mitchell et al., 2011a, MacDorman and Chattopadhyay, 2016], defined by perceptual mismatches. There is strong support for the occurrence of uncanny valley under perceptual mismatch conditions [Kätsyri et al., 2015]. This support stems from studies that have compared mismatched conditions of appearance and motion or face and voice mismatches. Specifically, the uncanny has appeared to emerge when more realistic modes (for example, high face realism) are paired with less realistic modes (for example, low voice realism) [Wang et al., 2015]. Given the relationship between facial features, voice features and emotional expressions, this thesis identified with the assumptions of the mismatch hypothesis.

Higher levels of negative valence and arousal have been observed in people experiencing the uncanny valley effect, where virtual character faces can be shown to increase negative valence compared to human faces [Cheetham et al., 2015]. In other words, this study found that avatar faces elicited the experience of a more negative emotional state compared to real human faces, and also categorically ambiguous faces. Furthermore, a contrasting perspective on these empirical claims found that other uncanny characters can be linked to experiences of amusement ahead of negative valence emotions [Mäkäräinen et al., 2015]. In this sense, there is a possibility that there is a positive valence to an experience of the uncanny, in that the weirdness or strangeness that is associated with the perception of the uncanny may stem from a place of entertained disbelief. A relevant point here comes from the famous french philosopher Henri Bergson who, in his text Laughter: essay on the meaning of the comic, notes that a human acting in a mechanical fashion is important for evoking a sense of amusement

[Bergson et al., 2018, Wilkie and Diddams, 2021].

This ambiguity was some motivation for our focus on valence in this thesis, which sought clarity on this important issue of classifying the emotional experience of the uncanny valley. The absence of emotional expressiveness has been measured to heighten perceived eeriness of digital characters in face-only stimuli presentations [Tinwell et al., 2011], and so the inclusion of positive and negative valence was also intended to help define a scale of eeriness at emotional dimensions beyond neutrality. Moreover, this work was intended to assess how people perceive emotional expressions in virtual characters.

A final note of import stems from the precise metrics required to measure the uncanny valley. It has been posited by experiment design reviews that any uncanny valley study that seeks to prove its existence must include emotional response metrics, such as affinity or warmth, as well as scales for eeriness and realistic human-likeness [Lay et al., 2016]. Moreover, multiple efforts have been made to adapt the indices which measure the UV [Bartneck et al., 2007, Ho and MacDorman, 2010b, Ho and MacDorman, 2017]. The need for these revisions has stemmed from two specific problems.

Firstly, given the relative recency of the UV phenomenon's empirical observation, strict definition of identifying terms has been a slow process. However, this problem has been the subject of iteration and development to the point of convergence upon operationalized commonalities. These commonalities are proposed to address the dimensions of the uncanny posed at the beginning of this section. Namely, direct measurements of eeriness, as opposed to strangeness and unnaturalness, are proposed to have the largest effects across studies. Similarly, familiarity has been commonly used as a dimension for human-likeness with high effect sizes, but the term may generate confounds in experimental formats with multiple stimulus exposures. Hence, realism is proposed as an appropriate measurement for human-likeness [Diel et al., 2022].

Secondly, as mentioned above, there are ambiguities as to the emotional character of the UV itself, and whether it is a negative experience or a positively valenced one [Mäkäräinen et al., 2015, Cheetham et al., 2015, Diel et al., 2022]. As outlined above, since mechanical human likenesses may be posited as a source of comedic amusement, there is a possibility that the experience of the Uncanny Valley is not captured by the strange, eerie, negatively valenced experience that is so often the focus. This may also be a methodological constraint, as many studies do not include specific emotionally valenced characters for Uncanny measurements. Furthermore, it has been established that emotion is easier to recognize in human faces compared to virtual ones [Beer et al., 2015].

In sum, the aspects of visual realism and vocal realism are richly entwined with the history

of the Uncanny Valley. Under the assumptions of contemporary cognitive science, our perception of vocal and visual realism must be considered as affordances; since, under the interface theory of perception, we perceive affordances rather than real objects, and since we construct instances of emotion based on the cues that the realism of these affordances can express, then what about this realism in virtual characters influences our empathic responses?

2.4 Methodologies to Assess Empathy for Virtual Characters

Emotional expressions in virtual characters, and their resulting perception by users, can be posited as a branch of 'affective computing', whereby emotions are elicited in users by the actions of the characters. As noted by [Dukes et al., 2021], the development of this domain of research is part of a general upswing in funding and interest in the affective sciences; in previous decades, the scientific focus on behaviour aimed at measuring behaviour while excluding cognition and emotion, before the turn towards measuring cognition which still excluded measuring emotion. In the modern scientific landscape, affective science is important for understanding the phenomenal experience of our day to day interactions, with each other, with ourselves, and with the technologies that are part of our lives, including synthetic agents such as chatbots, social robots and virtual characters.

For virtual characters in particular, empathy is often invited from a user by the emotional content portrayed in their behaviour, whether it is on screen in games or tv, as well as immersive 3D virtual reality [Pelachaud, 2013].

There is evidence to suggest that emotional expressions in computing systems have a greater tendency to influence the emotional valence of users as opposed to emotional arousal [Llanes-Jurado et al., 2024]. As such, we chose to mainly narrow our focus to the valence of the emotions (positive/negative) rather than arousal (high/low) in our efforts to understand and measure empathy for virtual characters.

2.4.1 Valence and Virtual Characters

In general, adding emotional elements to these systems have been proposed to improve user engagement and satisfaction with virtual characters as well as their overall believability [Maffiolo and Ochs, 2013].

Since it has been shown that a digital character verbally and facially expressing a positive or negative emotional state is sufficient to impact the emotional state of a human conversational partner [Qu et al., 2014], our work has sought to explore how manipulations of character face realism and voice fidelity could influence the empathic responses to emotions, as well as the ability to readily recognise such emotional states. This focus on emotional valence as expressed in faces and voices utilised the FACs powered face rigs of virtual characters as well as expressive acts of vocalisation, where the characters are describing an emotional mental state [Maffiolo and Ochs, 2013].

In terms of specifically positive or negative experiences, the valence of emotions has been investigated in experiments of avatar-avatar social interaction, where enhanced smiling salience has been shown to increase perceived positive valence when using said characters [Bailenson et al., 2016]. Such work has also been conducted on human-agent interactions, where higher realism of agents is observed to correlate with decreased negative valence [Volante et al., 2016], until the point where the aforementioned Uncanny Valley begins to emerge from the data. Moreover, recent prominence of virtual humans in social media has been studied, with findings suggesting a perceptual preference for positive rather than negative valence interactions [Ham et al., 2024].

In light of the incompleteness of any theory which has been developed to elucidate the relationship between virtual characters and emotional expressions, numerous works have explored their mutual relevance. For example, appearance and motion variability in virtual characters has been linked to changes in emotional reactivity [Mousas et al., 2018], while lower degrees of emotion expression in virtual characters have been correlated with perceptions of the uncanny [Tinwell et al., 2011]. Generally, the addition of emotional valence to a virtual character has been shown to increase its human-likeness [Slater and Burden, 2009]. This thesis proposes that this is due to the activation of empathy as a basic perceptual process that is most commonly experienced between humans in shared affect scenarios [Kiverstein, 2015]. In the use of virtual characters, there are numerous applications of emotion research that can be useful to understanding our interactions with human-like agents, and how these interactions are influenced by realism. In these interactions, it is proposed by this thesis that there is an observable experience of empathy in users.

One primary topic of consideration has been the use of avatars in emotive virtual contexts. Mood inductions, mediated by PC monitors, have been associated with higher senses of immersion and presence in virtual environments compared with neutral non-emotive conditions [Baños et al., 2004]. Emotion has also been prescribed as crucial for achieving a sense of presence in interactions with virtual-self representations [Ratan, 2012]. Emotion recognition between avatar user studies in video-conferencing contexts [Bailenson et al., 2006], have described the effects of behavioural realism on valence perception, and encourage future work to use more realistic avatar representations to further explore this domain. Studies in immersive virtual environments have also indicated that there are correlations in avatar attractiveness and reported friendliness [Yee and Bailenson, 2009]. As a metric, attractiveness has been used to assess positive emotional reactions to virtual agents as well as avatars, (See Figure 2.17) as previously developed by Ho and MacDorman [Ho and MacDorman, 2010a]. Avatar communication research has demonstrated the antithetical effect of negative valence and aggressive

reactions [Peña et al., 2009b] but these kinds of results have not been subject to full replication in computer-mediated dyadic conversations [Van Der Heide et al., 2013].



Figure 2.17: Some of the animated characters used by Ho & MacDorman to develop metrics of the Uncanny Valley [Ho and MacDorman, 2010a]

In terms of virtual humans, simple abstract emotive expressions by agents are not sufficient to influence emotion recognition processes in humans in screen-mediated avatar interaction, with a necessity to combine high fidelity emotional content with form and behaviour realism [Bailenson et al., 2006] which are both becoming more easily replicable with the development of photorealism in virtual characters. Stimulus fidelity that aims to replicate visual or other forms of modal realism has previously been defined as logically orthogonal to 'immersion', [Skarbez et al., 2017], a concept which in turn has proven to be both a significant subject of cognitive science research [Psotka and Davison, 1993] and measured to be influenced by emotionally valenced scenarios [Baños et al., 2004]. These theoretical links provided us with methods to investigate alternate effects of stimulus fidelity (i.e. realism) in emotionally valenced scenarios on human perceptual processes.

2.4.2 Empathy in Research

In general, emotion perception involves the capacity for empathy, which is normally measured along both cognitive and affective dimensions, apart from the common interpretation of empathy as a form of compassion or sympathy [Powell and Roberts, 2017]. Empathy in general can be described as a method for accessing the mental state of another via their actions and the affordances offered to the empathiser [Kiverstein, 2015]. Within this scope, cognitive empathy is often used to describe the perception of emotional thoughts, and affective empathy denoting a contagion of emotion between an expresser and a perceiver [Duan, 1996].

Empathy can be considered as both a contextual state and an individual trait [Powell and Roberts, 2017]. In computer based application, empathy has been a focus for studies on collaborative tasks with chatbot systems [Raamkumar and Yang, 2023]

and social robots [Pütten et al., 2014, de Kervenoael et al., 2020]. Moreover, trait empathy has been found to influence the aggressive treatment of video game characters [Gao et al., 2017], while state empathy has been measured in training simulations for doctors treating virtual patients [Kleinsmith et al., 2015, Krystallidou et al., 2018].

An additional form of empathy described in the literature is that of motor empathy [Jolien Van der Graaff and Branje, 2016], which can be expressed in the form of mimicry [Bavelas et al., 1986]. Unconscious mimicking of facial expressions of emotion has previously been recorded to differ between high and low empathy subjects [Sonnby–Borgström, 2002], and can be more pronounced for Zygomaticus major activity in dynamic stimuli compared to static ones [Rymarczyk et al., 2016]. Some models suggest that the actions involved in mimicking facial expressions may facilitate other aspects of empathy [Prochazkova and Kret, 2017].

The present thesis was partly motivated by the promise of examining whether the Uncanny Valley could be observed to influence motor mimicry [MacDorman et al., 2013]. In our final experiment, we do observe motor mimicry for emotional expressions in realistic virtual characters, that does not appear to be modulated by the uncanny valley. Instead, higher realism appeared to suggest a higher level of motor mimicry for positive emotions. As such, realism in terms of real human vocal expressions and high fidelity facial expressions, can be considered to influence mimicry, which in turn suggests that the affordances provided by realism encourage the process of adopting and impersonating the emotions of a virtual character.

An important addition is the notion that we can react to virtual characters as if they were real when it comes to empathic scenarios. This claim may be deduced from the findings of experiments which have presented virtual characters expressing pain [Slater et al., 2006a, Cheetham et al., 2009, Bouchard et al., 2013]. Specifically, these works have found a general tendency to react to virtual character pain as if it were real pain [Slater et al., 2006a], as well as relevant empathic traits associated with neural activity which indicates pain empathy for those same characters [Cheetham et al., 2009]. Indeed, there is an even further claim to be made that the 'familiarity' of these characters plays a role in empathy during expressions of pain, with evidence suggesting an empathic preference for known virtual characters with familiar voices, compared to less familiar characters and modified voices [Bouchard et al., 2013].

2.4.3 Physiological Correlates of Empathy

This work used two physiological measurements in an attempt to evaluate empathy for characters with different levels of visual and vocal realism. In order to objec-

tively measure emotional responses and empathy, we focused on two main techniques: electromyography and eye tracking. These methods have previously been used to assess emotional reactions to virtual faces, and their relationship to the Uncanny Valley [Cheetham et al., 2015, Mäkäräinen et al., 2015, Ciechanowski et al., 2019]. In employing both of these methodologies, it was important to thoroughly understand their general application as physiological measurements for the purpose of evaluating emotional responses.

Surface electromyography (EMG) is a technique for acquiring electrical activity naturally generated during movements, that can be useful at varying levels of activation and can be acquired by applying electrodes onto the skin so that they can read the activity from skeletal muscle tissue [Reaz et al., 2006]. Specifically, the units of activity generated during surface electromyography are voltage signals known as Motor Unit Action Potentials. This activity is acquired through the skin-based electrodes that measure analog voltage (i.e. the physical signal of the muscle activity, measured in volts) which are then converted to digital signals in the form of numerical units, using computational algorithms which are designed to decode the analog signals into Motor Unit Action Potentials. This digitalisation is conducted through the process of sampling, whereby the number of Hertz refers to the number of samples acquired each second. For example, surface electromyography conducted at 2000 samples of voltage signal per second is said to have a sampling frequency of 2kHz [De Luca, 2003]. While surface electromyography can be employed to measure the activity of many muscles in a body, it has long been used to link emotion to the activation of particular muscles in the face [Dlmberg, 1982, Lundqvist, 1995]. In particular, there has been considerable work linking the Zygomaticus Major (the cheekbone muscle that draws the mouth's angle upward and outward, therefore mainly responsible for smiling) to experiences of positive valence emotions, normally happiness, with the Corrugator Supercilii (the muscle that draws the eyebrows together, responsible for frowning) empirically verified to increase in activity under negatively valenced emotion conditions, usually in terms of anger, but also under conditions of sadness [Larsen et al., 2003, Sato et al., 2008] (See Figure 2.18).

Facial Electromyography has been an important tool in the domain of empathy research on face processing. It has already been established that higher trait empathy is correlated with increased zygomaticus major activity and corrugator supercilii activity in the comprehension of positively or negatively valenced emotional faces [Balconi and Canavesio, 2016]. Moreover, the activation of these muscles under other conditions of positive or negative emotional valence has also been linked to higher trait empathy [Dimberg et al., 2011]. The aim of employing facial EMG for the final experiment in this thesis was to consolidate the previous works which found evidence for

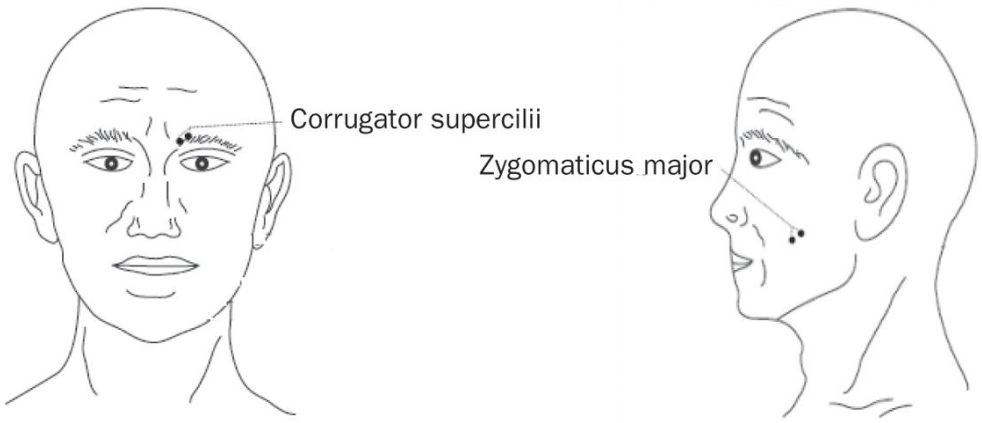


Figure 2.18: Corrugator supercilii and Zygomaticus major face muscles that are commonly associated with negative and positive emotional valence respectively [Martínez-Fiestas et al., 2015]

subjectivity in emotion perception at varying degrees of character realism. This consolidation was proposed as a method to better understand empathy for virtual characters through measuring the activity of facial muscles during emotional scenarios.

Our second physiological measurement was conducted with the use of eye-tracking; we were interested to see how participants gaze fixated on virtual characters that expressed emotions through vocal utterances and facial movements. In general, eye tracking can be performed non-invasively by modern video-based tracking systems. These systems use infrared light to illuminate the pupil and reflection on the cornea of a viewer. The tracker is then able to record the movements of the eye using a camera, while also plotting raw eye gaze coordinates by calculating x and y positioning within the specified range of the camera. These coordinates can be used to measure pupillary responses to stimuli, saccade movements, and various fixation metrics [Stuart, 2022]. In emotion recognition tasks, raw eye tracking coordinate data (i.e. where in a given area of interest an eye rests or moves to and from) are used to calculate factors such as fixation duration and quantity [Tarnowski et al., 2020]. These were the two elements focused on in our work. A fixation is normally defined as a brief moment where the eye is stationary within a given area of interest, normally measured as lasting between 150 and 600ms. Fixation metrics can differ, with various methods utilizing measurements for the first fixation within an area of interest, the amount of fixations in a given area, the frequency of fixations and the mean duration of fixations across an area [Cummins, 2017]. Since fixations may be used to measure a viewer's spatial focus over time [McDonnell et al., 2009], this work employed their use to measure attention to emotional stimuli. Importantly, this method is comparable to the Facial EMG method, in that an eye tracker also operates at a

specific sampling frequency, and as such is designed to capture x amount of imaging samples per second at a prespecified Hertz rate.

In addition, eye tracking methods have been employed to study emotional expressions in faces such as ours [Schurgin et al., 2014]. In particular, the work of [Schurgin et al., 2014] highlights visual responses to 'basic' emotional expressions [Ekman and Friesen, 1978]. Once again, there is an opportunity to couch the cues provided by these expressions in the language of affordances, with a main assumption of this work stating that, when users eyes fixate on the eyes or mouth of an emotional virtual character, they are not fixating on a presentation of a basic emotion, but rather a collection of cues that allow for predictive neural activation strategies to construct the occurrence of an emotion [Gendron and Barrett, 2018].

Eye-tracking methodologies have also been used in studies which aimed to elucidate the influence of virtual agent body positions on emotion communication [Marschner et al., 2015] as well in medical training simulations [Volonte et al., 2019], and have previously been combined with electromyography in experiments on emotional expressions in virtual characters [Joyal et al., 2014]. Other work using both of these physiological measures has observed that children with ASD respond more readily to emotions in avatar faces, particularly expressing negative valence, compared to real human faces [Pino et al., 2021]. As with facial electromyography, data on eye-fixation for human faces have been associated with trait empathy [Cowan et al., 2014]. In light of these works, our final study aimed to use both gaze fixation and facial EMG data to understand whether certain conditions of realism were more likely to provoke changes in psycho-physiological responses among healthy adults.

2.4.4 Summary

The background literature described in this chapter has informed the main thrust of this thesis, which asserts that we can experience empathy for virtual characters at varying degrees of realism, and that the voices of these characters, whether absent, synthetic or real, determine the degree of cognitive empathy that we have for them during emotional scenarios. While these are the main findings derived from our experimental work, the most general finding from across all of our experimental results is that the combination of voices and faces in virtual characters is the aspect of their presentation that most affords emotion perception, and empathy, compared to voice and facial appearance in isolation. To conclude this section, there are several important points to reiterate.

Virtual Characters

Virtual characters can come in different forms, from embodied avatars to independent agents. The difference between these two forms of virtual characters is defined by control, but the appearance and vocal characteristics of interactions with animated humans are consistently efficacious regardless of this element of control. As such, this thesis covers both the human controlled virtual form and the model that is subject to such control, and develops a theoretical approach to empathy which is unique to the study of virtual characters.

Perception of Realism

This theoretical approach is based on the subjectivity of perceiving affordances in realistic characters. These characteristics of voice and appearance are impossible to separate from the notion of realism. Since virtual characters, within the scope of this study, are representations of real people, the degree to which their qualities are perceived as real has a direct relationship to their affordances for emotion perception. Additionally, the misalignment of these affordances can produce a reaction of discomfort or strangeness in users, in a manner which is highlighted by the theory of the Uncanny Valley.

Direct Affordances

These affordances are perceived directly, as experiences of understanding, feeling and sympathising with the emotions of virtual characters. In our studies, these characters express emotions via vocal and facial expression behaviour. Since these are prominent methods of emotion communication between human agents, people tend to react as if there is a real person present in front of them when they are simulated by a virtual character. This thesis draws upon the contemporary literature in cognitive science to claim that this reaction is caused by a direct relationship to emotional stimuli, whereby we construct an emotion based on affordances that have been learned prior to an interaction. The emotions are not built upon a preexisting mental model of emotion, so much as they are constructed with a learned action, which is the skill of emotion perception. This bottom-up approach to perception, as applied to virtual characters, requires a novel combination of vocabularies from computer graphics and cognitive science, in order to describe empathic responses to such characters.

Empathic Responses to Emotional Affordances

In the following chapters, these affordances are gradually narrowed down. From studies using avatars in screen-based and immersive contexts, it is established that there is

some effect of having realistic human representations, and concurrent voice stimuli, on the perception of emotion in virtual characters. Additionally, this effect of human-like representation may be modulated by individual user traits.

Following from these findings, it was established that variations in voice or appearance realism alone differ very little by way of providing affordances for emotion perception, as of course no emotion is truly unimodal.

In contrast, results indicate much more variation by way of emotion intensity when highly realistic voices are combined with highly realistic appearances, compared to lower fidelity counterparts.

To whit, these combinations are not only important for emotion perception, their combination appears to be an affordance by itself that allows for the construction of an emotion. As such, combining these qualities in a virtual character allows for empathic responses regardless of the degree of realism. It appears that this realism does not have a strong effect on physiological signs of empathy, so much as virtual characters themselves seem more suited to afford positive emotion perception ahead of negative. Moreover, while reports of affective and compassionate dimensions of empathy seem to be influenced by visual realism, cognitive empathy is measured to be significantly stronger for real human voices, when they are present in virtual characters. The following sections outline the series of experimental works which concluded with this novel finding, beginning with the study of trait influences on avatar embodiment.

3 Experiments 1A-C: Trait Effects on Experiences of Humanoid Avatars

The following works focus on user experiences of real-time motion captured 3D virtual avatars. The first two studies were conducted in 2D video-conference settings, with the third experiment expanding this work to study the effect of human and non-human appearances on the perception of avatars in a 3D virtual environment, with a view to expanding the findings from the first two experiments on how subjective user traits influence this perception. Our experiments were designed with the intention of exploring previously defined perceptual illusions that occur with avatar-use in virtual reality.

Moreover, research described here was aimed at empirically evaluating four separate dimensions of human-avatar interaction. The first was humans-as-avatars, with experimental conditions that were designed to measure changes to subjective perceptions of self-face ownership and self-concept. This dimension was measured in dyadic personalised settings for experiments 1A and 1B, with experimenter and participant both embodied as avatars. For experiment 1C, participants by themselves experienced generic and non-human avatar embodiment, to assess which external cues, specifically priming and character appearance, could influence virtual avatar perception.

The second focus was other-perception, with the unique design of the studies 1a and 1b outlined below among the first to measure social presence in a video-call between two human-driven avatars.

The third emphasis was on the experiential content involved in avatar use, as there were measurements for emotion induction, fatigue and behaviour change included in the data collection.

Finally, these experiments showed how the self-reported perception of these avatars, both in 2D and 3D contexts, was subject to the individual differences intrinsic to each

user. This difference in subjective experience combined with reports of embodiment is considered to be related to how concepts become salient to individuals of differing dispositions.

3.1 Experiment 1A: Pilot Study of Emotion Communication

For the pilot study, one experimenter interviewed individual participants who were embodied in self-representational avatars. It was designed to test our hypotheses around whether participants could experience senses of ownership and social presence, and how mood induction and self-concept varied between conditions. This experiment was designed to explore a broad perspective on the experiences of emotion with virtual characters.

Fourteen participants were recruited to take part in this experiment. They were provided with information and consent forms prior to the experiment. The participants were recruited using university mailing lists, and comprised of undergraduate students, postgraduate students and research staff. The mean age was 26.6 years, and there was a reasonably representative spread of 7 males, 6 females and 1 non-binary participant. Ethical approval was provided by the university in advance of this recruitment and all participants signed a consent form prior to the experiment. They were each entered into a draw for a €50 voucher for their efforts.

3.1.1 Stimuli Creation

Zoom platform was used to run the conference call, chosen for its ease of use and free distribution qualities. The participants used a modern avatar application with accurate tracking of 198 facial landmarks and moderately rich facial animation of 51 expression blendshapes[1] which included a Zoom camera plugin. The application was chosen as for allowing the virtual face to directly map the visual and kinesthetic actions of each participants face and head as in the study by Kokkinara et al. [Kokkinara and McDonnell, 2015]. The faces of the respective avatars were displayed side-by-side in gallery view on the video call.

Both the experimenter and participant embodied their chosen avatars from the beginning of the call. The cartoon-like appearance of individual avatars was predicted to modulate how the participants perceived themselves and the experimenter. Specifically, it was

[1] https://www.hyprsense.com/hyprface



Figure 3.1: Pilot Experiment: Some examples of the customised avatars created by the participants.

relevant that the avatars possessed larger eyes and also expressed enhanced smiling features compared to real human faces, in a manner which was considered to afford positive emotional valence. Such smile-enhancement features have previously been implicated in improving positive affect induction in virtual avatar faces [Bailenson et al., 2016]. The experimenter embodied the same generic avatar for each experiment with a view to consistency within conditions. One of the limitations inflicted on this study by the COVID era was the necessity for participants to use their own personal computers, webcams and internet connections to complete the experiment tasks. However this drawback was tolerable, given that, since our participants would have been experiencing multiple Zoom calls a day, it was more appropriate to try and measure these effects from the place where they usually conducted the same activity. In this sense there was also a beneficial influence on the integrity of the study conducted.

Prior to the experiment, the participants personally designed their avatars to represent themselves. There were a variety of detailed parameters to change in the character design, with functions for shaping face parts and shading tones which the participants adjusted to closely represent themselves.

Each participant was individually trained on operating the avatar software. It helped to gauge how effectively the face-tracking software was able to integrate into the individual operating systems that the participants were using. There was no way to control for this extraneous factor given the circumstances, but a sufficient standard of interaction quality was attained for all subjects prior to the main experimental task phase.

Participants were individually welcomed into a pre-arrange Zoom conference call and asked to engage in a video-conversation using a personalised avatar in their likeness for the entire experiment.

The conversations took place in two separate blocks following two separate Mood Induction Procedures (MIPs). Such procedures are used to elicit specific emotional states in experimental participants, using different kinds of methods. Mood inductions may be achieved using the free mental generation of ideas, whereby the intended emotional state of the experimenter is activated mentally by the subject. For example, subjects may be hypnotized or encouraged to ideate. A second type of MIP can be an imaginative state that is guided, such as when an ideation is combined with music of a specific emotional tone (as used in Experiment 1B.) For the present pilot experiment, participants were subjected to the third category of MIP, whereby they were exposed to emotion inducing material [Gerrards-Hesse et al., 1994]. The use of MIPs has been commonplace in avatar-interaction experiments in immersive VR [Baños et al., 2004]. In the context of our experiment, deploying this technique meant using stimuli in the form of positive and negative newsreel clips with appropriately valenced audio backing in each of the two videos.

The first video[2] was chosen to elicit positive emotional content such as joy, amusement, respect and charm. This contrasted with the second video[3], which is presented in a negative light and intended to put the participants in a state of unease and discomfort. See Supplemental movie for an example of the experiment.

After each of the two videos, the participant and experimenter engaged in four minutes of conversation, in order to encourage the participant to fully engage in the social use of their avatar using pre-specified neutrally worded dialogue prompts.

It was predicted that the cartoon-like appearance of the avatars was likely to make interactions difficult during the negatively valenced conversation. Previous work with avatars has described difficulties in attempting to interpret wide-eyed expressions of negative valence [Koda and Ishida, 2006], while previous work which outlines the effects of positive enhancements of avatar features [Bailenson et al., 2016] also motivated this prediction.

[2]https://youtu.be/WFWQbTf40Ik
[3]https://youtu.be/z7d4P3ybcXc

3.1.2 Experiment Design

Following the two videos and two conversations, the participants were asked to report on a subjective questionnaire on the topics outlined below (see Table 3.1). The metrics used were derived from a body of similar research, using methods adapted from established evaluations of social presence, enfacement and embodiment in avatar interaction settings.The responses included a commentary section which we analysed separately to the rest of the questionnaire data. The structure of this experiment was adapted from previous efforts to standardise data gathering methods in emotional elicitation experiments [Kossaifi et al., 2019]. The experiment was conducted with participants and the same male experimenter for all fourteen recorded versions.

We were constrained somewhat in our valence examination, given that we chose to distribute only one questionnaire at the end of the experiment, instead of one after each valence stimulus. We decided that the data we sought to derive from the questionnaire could be confounded in the second instance. If a participant was asked, for example, whether the videos affected their mood, they might be more cognizant of that information for the second video and questionnaire. The main Experiment (Section 4) was designed to account for such demand characteristics.

Another measurement constraint impacted attempts to control for animation and face-tracking accuracy. In order to ensure a fair comparison across the board, this quality was post-rated following each experiment recording. Subjective scores for an objective quality such as face-tracking accuracy was considered to be a potential confound, so a post-experiment analysis rated the quality as either high or low.



Figure 3.2: Video call between participant (left) and experimenter (right), both embodied in virtual avatars and viewed side-by-side on-screen.

The subjective questionnaire we used to obtain data from this experiment was constructed using a body of literature in the area of human-avatar interaction. The questions we used which sought to elucidate subjective senses of enfacement (see Table 3.1), are adapted from previous research which attempted to standardise the categories of

information-seeking queries that should be included in these kinds of experiments. We also used the Linguistic Inquiry and Word Count to analyse data from an open-ended comment section at the end of the questionnaire.

Table 3.1: Pilot Experiment: Questions arranged by group and variable name. Each question could be answered on a scale from 1 – "Not at all" to 7 – "Extremely".

Group	Variable Name	Question
Emotional Valence	Affect Level	*"The videos affected my mood."*
	Positive Expression	*"I found it easy to express my emotions after the positive video."*
	Negative Expression	*"It was difficult to express my emotions after the negative video."*
	Perceived Positive Affect	*"After the positive video, I felt more friendly than usual."*
	Perceived Negative Affect	*"After the negative video, I was less talkative ."*
	Valence Impact	*"I found it more awkward using my avatar after the negative video ."*
Enfacement	Self-Identification	*"I felt the sensation of being my avatar"*
	Agency	*"I felt that I could control the virtual face as if it were my own face."*
	Ownership	*"I felt as if the avatar's face was my face."*
Embodiment	Motor Control	*"I felt that my movements influenced the movements of the avatar"*
	Lack of Motor Control	*" I felt limited by my avatar's lack of movement"*
	Disembodiment	*"I found it difficult to talk while using my avatar"*
Proteus Effect	Avatar Realism	*"I felt that my avatar was similar in appearance to me"*
	Proteus 1	*"During the experiment, I was more talkative than usual."*
	Proteus 2	*"I felt more attractive as my avatar"*
Social Presence	Item 1	*"I felt like I was in the same room as the other person"*
	Item 2	*"The other person seemed stimulated by the conversation."*
	Item 3	*"I felt that I mimicked the movements of the other avatar."*
	Item 4	*"I felt as if the other person was listening when I spoke"*
	Item 5	*"Our conversation felt like a face to face meeting."*
	Item 6	*'I found the other face trustworthy"*
	Item 7	*"I perceived empathy in the other face"*

Subjective Questionnaire The valence items were extracted from studies that made efforts to standardise emotional appraisal questions [Roseman, 1996] and their use in avatar experiments similar to ours [Bailenson et al., 2006], and were adapted to suit this investigation. The questionnaire items used to measure enfacement have been extracted from questionnaires used in research into self-face recognition and the SoE [Gonzalez-Franco and Peck, 2018, Panagiotopoulou et al., 2017, Sforza et al., 2010]. These items have been used for subjective reports of face ownership and have been considered suitable for avatar-interaction studies [Suh et al., 2011].

The question items for Proteus measurements were composed based on the self-

perception theory [Yee and Bailenson, 2007] which underpins the occurrence of transformed self-representation, and work which has linked avatar appearances to properties of extroversion [Koda and Oguri, 2019].

Specifically, [Yee and Bailenson, 2007] posits the example of measuring friendliness after inhabiting an attractive avatar. Our item *Proteus 2* in Table 3.1 reflect the converse inquiry, with perceived attractiveness measured based on the friendly avatar facial details described in 3.2. Item *Proteus 1* is similarly composed to capture a self-reported change in behaviour. Talkativeness has been described as a Proteus measure for inhabiting animated avatars of improved appearance [Koda and Oguri, 2019].

The social presence items were extracted from previous study [Nowak and Biocca, 2003] and adapted for the purpose of investigations such as ours. The realism and social presence determinants are derived from more exploratory research in the area. Social presence is a topic that has received much attention from the VR research community, as it is a strong indicator of a user's level of immersion in a social VR or AR platform, and has been used to measure the degree of 'success' of a virtual social interaction [C.S. Oh and Welch, 2018]. Presence is also considered to be bound inexorably to perceived SoE [Mennecke et al., 2010]. As such, there has been an extensive effort to standardise the acquisition of information from subjective questioning on the topic.

Linguistic Data The Linguistic Inquiry and Word Count (LIWC) provided us with the means to compare comment data to the questionnaire data in an attempt to quantify the data we were looking for. As described below, the LIWC was used to analyse participants open ended responses to the content. Specifically, they were asked "Please describe, in at least five sentences, your experience of using the avatar for conversation. Focus on how you felt during the interaction. Please be as detailed as possible" [C.S. Oh and Welch, 2018].

The intended purpose was to extract implicit and explicit cues that consolidated evidence for the effects that the questionnaire data would provide. While we were interested to see if our embodiment and social presence questions could potentially be reflected in this data, the LIWC is considered a valid indicator for the content of emotional expression [Kahn et al., 2007], and we hoped that it would shed light on the effect of the valence conditions on avatar interactions.

We anticipated that participants using personalized avatars for dyadic video conversations would feel ownership of their embodied self-avatar (Ha), report social presence with an other-avatar (Hb), find it easier to discuss positive content than negative content during the experiments (Hc) and experience some alteration of perceived self-concept (Hd).

3.1.3 Results

Each questionnaire item was analyzed separately. There were different kinds of information we were hoping to get from the results. We conducted a one-way Analysis Of Variance (ANOVA) with between–subject factor Quality (low, high) for the dependent variable (subjective responses). We also conducted a Spearman's Rank Order Correlation test to compare the results of some of the questions. The comment data, also analysed individually, was processed using the Linguistic Inquiry and Word Count, in order to provide evidence for the effects that we sought to explicate in conjunction with the variance analysis. The data from the LIWC was taken at face value, although we used a repeated measures ANOVA to analyse the linguistic affect data.

Since most of the participant questionnaire data was collected via a 7-Point Likert-Scale, a score of > 4 is assumed to indicate an affirmation of effect, and a score of < 4 would indicate an absence of effect. The analysis for these results was conducted using a t-test of means against the constant of 4.

There were several null and alternative hypotheses to contend with in reviewing this dataset. We predicted that positively valenced stimuli would produce a beneficial effect on the avatar use, and that the negative stimuli would make them more difficult to use. We also predicted that senses of face and body ownership would be evident post avatar interaction. Our other alternative hypothesis was that there would be an obvious perception of social presence in the data, despite not being able to see one another on screen, and only interacting using the avatars.

The ANOVA was justifiable given the distribution of age, gender and ethnicity of the set of participants. They also did not interact with each other and as such we felt that the independence assumption was appropriately preserved despite the relatively small population tested.

Emotional Valence

Our t-test results provided evidence for an interaction between the valence stimuli and the moods of participants, represented by the first item on Table 1 for Affect Level ($M = 4.57, SD = 1.82$). As the blue bars in Figure 4.6 depict, the impact of positive valence is more prevalent in the results than the effect of negative valence. This is evidenced by the measurements taken for the Positive Expression ($M = 5.43, SD = 1.28, t(14) = 4.16, p = 0.001$). However, there was no significant effect detected for the Perceived Positive Affect measure ($M = 4.21, SD = 1.76 t(14) = 0.45, p = 0.65$). There is a contrast here with the results for the other question items on negative valence,

with the strongest influence being observed in the Negative Expression question data ($M = 3.07, SD = 1.35$), meaning that it was not difficult to express emotions after the negative video ($t(14) = -2.32, p = 0.036$). The other values for the negative valence questions provide even lower indications of negative affect, which we considered to be evidence for an effect of non-negativity, when combined with data from the LIWC (see Section 3.1.3). Implications and causes for this are discussed below.

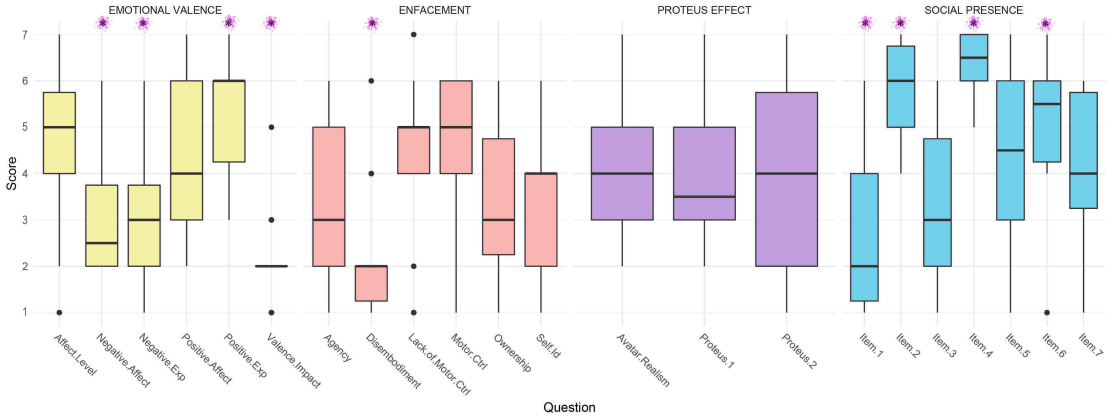


Figure 3.3: Pilot Experiment: Results for questions Emotional Valence, Enfacement, Proteus Effect and Social Presence. X-Axis labels are the variable names from Table 3.1. Star indicates statistical significance < 0.05 against the reference value '4'.

Enfacement

The results for face and body ownership questions between subjects indicated that senses of enfacement and embodiment were not felt by the participants. There is an observable effect for other question groups, but it is less pronounced for the embodiment scores (see Figure 4.6, grey). The highest means in the data gathered on SoE come from the Motor Control ($M = 4.64, SD = 1.65$) and Lack of Motor Control ($M = 4.36, SD = 1.82$) measures. These results are consistent with previous studies on the topic, and confirm the importance of movement for SoE. The face ownership results are harder to interpret, as the responses for ownership ($M = 3.43, SD = 1.76$), and the face-control measure that intended to augment the data for enfacement ($M = 3.57, SD = 1.80$), yielded results insignificant results. This suggests that there were only weak effects for face and body ownership. Potential reasons for this are explicated in our discussion.

Proteus Effect

In order to determine if there was evidence of the Proteus Effect, we conducted a non–parametric Spearman's Rank Test to examine correlations between the results on enhanced feelings of attractiveness and ability to talk easily, friendliness, awkwardness,

and talkative ratings. These measures were informed by studies conducted in immersive VR conditions, which suggest that attractiveness and positive feelings of friendliness and extroversion [Yee and Bailenson, 2009], and constituted our empirical search for evidence the Proteus effect. There were no findings in the Spearman's results to indicate a main effect. Interestingly, previous studies that attempt to quantify this effect in computer-mediated dyadic avatar interaction have also failed to demonstrate this effect [Heide et al., 2013], and so our evidence seems indicative of an upper boundary or a necessary condition that is missed in-between immersive VR and dyadic remote conversation, which we discuss below.

It is also possible, since the overall experience was reported and described positively by participants, despite the presence of the negative stimulus, that the ebullient appearances of the avatars did demonstrate a level of the Proteus effect. Evidence for this can be found in our LIWC results on affect, but are only suggestive and not confirmatory of this hypothesis.

T-test results indicate that there was no main effect for the Proteus 2 item ($M = 3.71, SD = 2.23$) from the reference value of 4 ($t(14) = -0.47, p = 0.64$) indicating the participants did not feel significantly more talkative with their avatar (see purple bars in Figure 3.3). There was a similarly unconvincing result for the Proteus 2 item on attractiveness ($M = 4.07, SD = 1.49$) that reflected no major effect ($t(14) = 0.18, p = 0.86$). These results are in line with the information derived from the Spearman's correlation results.

Social Presence

The social presence scores were much more pronounced than the embodiment or Proteus scores. The conversational aspect of the design meant that each participant had to engage with using their avatar in a social context. This is reflected by the results of the social presence questions (see light blue bars in Figure 4.6). There was a noticeable effect on user engagement, with a recorded measure of social presence on Items 2-7 in Table 3.1. The questions where social presence was most visible were for Item 2 ($M = 5.79, SD = 0.94$) and Item 4 ($M = 6.29, SD = 0.79$). The t-test results for these items confirmed high levels of statistical significance, with Item 4 ($t(14) = 10.36, p = 0.00$) slightly stronger than Item 2 ($t(14) = 6.85, p = 0.000012$). Social Presence, by these measures, can be confirmed as an observed result. The participants also evaluated trust in the other avatar based on Item 6 t-test results ($M = 5.14, SD = 1.56, t(14) = 2.73, p = 0.017$). This is consistent with the appearance of these avatars, as larger eyes and face width have been associated with heightened

senses of trust [Ferstl and McDonnell, 2018]. Conversely, the participants did not perceive empathy in the avatar face ($M = 4.14, SD = 1.65 t(14) = 0.32 p = 0.75$).

Quality

A between-groups ANOVA was conducted to determine the effect of quality of facial tracking and we found a main effect of quality on attractiveness ratings ($F(1, 12) = 5.36, p < 0.04$), where participants felt more attractive as their avatar when the quality was high than when it was low. This result is consistent with other work in the field, particularly a study by Kokkinara et al. [Kokkinara and McDonnell, 2015] which found that lower quality facial tracking reduced the perceived appeal of the avatar in their reduced animation realism condition. It is also consistent with our finding that face animation quality did not impact agency or ownership under these conditions, although our study did not deliberately manipulate this factor. Another main effect of quality was found where participants found it more awkward using the avatar after the negative video when the facial tracking was poor ($F(1, 12) = 5.8, p < 0.04$).

Quality was measured in order to ascertain whether poor face-tracking and latency had an effect on the perception of avatar use that we intended to measure. Surprisingly, there was no other observable effect on the results despite some obvious differences in quality across participants. In particular, we predicted that the agency aspect of embodiment might have been impeded by poor function. However, this lack of effect may be explained by typical interaction quality with video calls, which often contains varying levels of latency quality without avatars. Also interesting to note is that animation quality had no effect on Ownership in the enfacement study by Kokkinara et al. [Kokkinara and McDonnell, 2015].

Linguistic Inquiry and Word Count

The LIWC allowed us to compare the questionnaire data with language use in the open ended comment format, with a word count of 1606 across the fourteen participants.

The effect size observed in the analysis of variance shows some similarity to the returns from the LIWC, which indicated an average of 2.74 'body' related words between subjects open ended comment sections, making up around 2% of the total word-count. Having not included bodies or the topic of embodiment in priming the participants, these results are indicative of a low level of SoE being achieved across the board, although significantly distant from a fully immersive embodied experience.

A useful observation can also be derived from the LIWC results on social presence. A

total of 7% of the total word count were identified as 'social' in content; relating to others via the self. This is a significantly larger proportion of the word count than the embodiment scores, and this difference is reflected in the questionnaire data as well.

A repeated measures ANOVA was conducted on the LIWC results for positive and negative affect terms. A main effect was discovered between the two variables that is consistent with the scores on the questionnaire data ($F(1, 13) = 34.33, p < 0.00007$). Positive affect scores ($M = 3.49$) were significantly stronger than negative affect scores ($M = 0.87$), which implies an overall positive experience despite the presence of the negative stimuli.

3.1.4 Discussion for Experiment 1A

These results revealed numerous implications for Hypotheses Ha-d.

Ratings for self-face recognition, ownership and agency were not high (Ha) in our Pilot, at least at the level we had predicted based on other studies on ownership and enfacement [Kokkinara and McDonnell, 2015, Porciello et al., 2018]. The most interesting results, for Motor Control and Lack of Motor Control measures (See Table 3.1), seemed to suggest a limited range of motion felt by participants, despite the fact that they felt that their own movements were depicted by the avatar. An explanation for this result could be derived from the fact that the avatar's movements did not represent the full range of human motion. The participants saw their own movements represented by the avatar's movement, but felt as if their avatar could not fully represent all of their implicit motor representations.

The Quality analysis also revealed no difference in these ratings for low or high quality, which indicated that the avatar use in this context was not sufficient to achieve levels of embodiment or enfacement, regardless of video quality.

The social presence scores were much more pronounced than the Enfacement or Proteus scores. The conversational aspect of the design meant that each participant had to engage with using their avatar in a social context. This is reflected by the results of the social presence questions (see blue bars in Figure 3.3). Evidence for social presence was observed on these measurements (Hb). The perceived stimulation and attentiveness of the other avatar could be due to the nonthreatening and engaging visage offered by the cartoon characters. In particular the evaluation of trustworthiness in the Pilot matches previous work related to avatar appearance [Ferstl and McDonnell, 2018]. At the very least, the use of the avatars was not measured to detract from the social presence you would expect in a regular face-to-face meeting.

We had predicted that the effect of the negative stimulus would be observable in the responses to the questionnaire items about the negative video. We expected that the avatar effect scores would not generally be high for post-negative conversation, based on previous experiments on lighting manipulation and emotional expression [Wisessing et al., 2020]. However, the positive affect scores suggest that the use of avatars for post-positive conversation displayed stronger mood induction efficacy (Hc), and that participants were more likely to feel 'friendly'. This result is consistent with similar studies [C.S. Oh and Welch, 2018].

We conducted a non–parametric Spearman's Rank Test to examine correlations between the results on enhanced feelings of attractiveness, friendliness, awkwardness, and talkative ratings, as a way of measuring self-concept change (Hd) and by extension a means to identify evidence for the Proteus Effect. There were no findings in the Spearman's results to indicate evidence for self-concept change. Interestingly, previous studies that attempt to quantify this effect in computer-mediated dyadic avatar interaction have also failed to demonstrate this effect, and so our evidence seems indicative of an upper boundary or a necessary condition that is missed in between immersive VR and dyadic remote conversation.

3.2 Experiment 1B: Main Experiment of Emotion Communication

From the previous results, it became evident that there is a rich domain of knowledge to be carved out here. Given recent rapid advancements in animation technology, we can and do expect that virtual humans will be represented in more and more realistic detail. As such, developing models of the effects such virtual characters, either in embodied or social interaction contexts, have on human perception and cognition is important for understanding the scope of potential effects and reactions in these kinds of scenarios. The second experiment was designed to test our developed hypotheses around the effects of dyadic conversation with personalised avatars, under different conditions of valence and video-feed content.

More precisely, two conditions were defined for the purpose of this experiment. Positive and negative emotion elicitation procedures were conducted for conditions of avatar-avatar conversation and avatar-voice conversation, where the experimenter did not engage using an avatar. This addition was made to use Avatar-Voice conversations as a benchmark to measure the impact of another avatar in Avatar-Avatar conversations, and whether the difference could cause variation in our measurements. The order was

evenly split between experiments, with half of participants experiencing non-avatar condition before the avatar-avatar condition. The display of the participant's on-screen self-representation remained consistently sized throughout all conversations. However, the voice-only condition was proposed as a method to compare emotionally valenced avatar scenarios with voice agent scenarios.

There were a number of other changes between this format and the pilot. Our data collection methods had greater depth, as the participants were asked to fill out four separate questionnaires. The question items were modified for application in the altered conditions, and there was no comment section included in the collection. Our hypotheses were similarly modified to reflect the results above.

Twenty participants were recruited separately via college mailing lists and international online distribution channels. Each participant provided consent forms, and were given information sheets prior to the experiment. One of the participants was a returning volunteer involved in the pilot study. This was considered permissible given the differences in the study design on terms of structure, measurements, task performance and avatar software, as well as the long period of time separating experiments, and the recruitment restraints imposed by multiple national lockdowns. Two experimenters (1 Male, 1 Female) tested this group of participants. The participants were split evenly on gender between experimenters, as there were 10 Male and 10 Female volunteers with a mean age of 25.6 years.

3.2.1 Stimuli Creation

The participants for this experiment used a different avatar application to those in the Pilot Experiment. The developers of the previous avatar application chose to close their Hypermeet venture, and so we took the opportunity to use similar software, with the added capability of a self-representational avatar generated from a photo, which saved requiring the participants to construct their own from scratch[4].

The software was slightly more significant in terms of system load, so only participants with higher quality machines, webcams and internet connections could take part. This coincidentally improved the experimental conditions and removed our need for grouping participants on the factor of quality. The experimenters used self-representational cartoon avatars, as the generic ones provided were qualitatively different in general appearance to the self-representational characters generated by a photograph. We wanted to avoid character-based confounds.

[4]https://loomai.com/



Figure 3.4: Main Experiment: Some examples of the customised avatars created by the participants using the photo avatar creator tool.

For consistency, each participant was asked to pick from one of four neutral office backgrounds in the avatar application. One of the participants used a blank background, as technical difficulties prevented full background loading. However, the avatar itself was fully functional for this instance.

Prior to the experiment, participants were also asked to configure several settings on Loomie software, such as activating Video Mode (allowing avatars to follow head movements of participants), activating Eye Tracking, importantly allowing the avatars to simulate elements of direct gaze [de Hamilton, 2016] and setting "Gesture Strength" at minimum as the Loomie avatars would otherwise perform hand gestures on their own.

3.2.2 Experiment Design

A within-subject design was adopted for the experiment, considering two independent variables: PartnerRepresentation with two levels: *Avatar* (with experimenters represented by a personalized avatar) and *Voice* (with experimenters' representation hidden and therefore only their voice provided to the participants), and Emotion with two levels *Positive* and *Negative* corresponding respectively to the positive or negative emotion supposed to be induced by the pieces of music. Note that the participant always used an avatar representation regardless of the experimenter representation. Moreover, when the avatar of the experimenter would be hidden, the size-on-screen of the participants' avatar would remain the same.

The *Voice* condition was chosen as a means to compare the *Avatar* conversation to a more recognisable form of tele-communication. The aim was to see whether the participant engaged more or less with their own avatar depending on the condition, and

also to assess whether emotional content was more salient in the presence of an avatar conversational partner on the screen. It was also predicted that there would be observable changes in participant feelings of Ownership as a result of increased self-avatar focus on screen.

There were two mood induction and conversation tasks for either PartnerRepresentation condition, making for a total of four tasks comprising of 90 second stimulus exposure and 120 second conversation periods for a total of eight minutes of avatar social interaction. These tasks were consistently performed in the same order, with the positive valence preceding the negative stimulus for each condition.

However, the PartnerRepresentation conditions were counterbalanced to eliminate any potential statistical biases that could arise from a strict task order hierarchy. Similarly, the musical pieces used for positive and negative valence were used at different task times and each video condition, evenly distributed between participants.

For the procedure, the participants were welcomed individually into a conference call over Zoom with an experimenter. The experiment started with a training phase, in which participants were asked to explore the full movement parameters of their avatar head, and present two emotional expressions that can be displayed on the face.

More precisely, participants were trained on the expressive use of the avatar; they were directed to make a happy face and a sad face using particular instructions (e.g. pull your outer lips down to make a frown). This helped in ensuring that the use of avatar for conversation would be conducive to emotion elicitation procedures.

After training, each participant engaged in two emotion induction tasks for two separate conditions. The emotion induction task used a modified MCI technique. This process involves music exposition and participant ideation of thoughts relating to the intended emotional measures.

This method is widely considered as an effective means of emotion elicitation [Eich et al., 2007], and the musical stimuli were not as widely pre-consumed compared to each of the clips used in the Pilot Experiment. The pieces were chosen from a wide range of verified elicitation pieces [Eich et al., 2007]: *Mozart Eine kleine Nachtmusik Allegro* and *Vivaldi Four Seasons: Spring III Allegro* for positive mood induction, *Schumann Träumerei* and *Chopin Prelude in E-Minor (op.28 no. 4)* for negative mood induction.While not sufficient to elicit positive or negative emotions in isolation, these pieces have been shown to be effective valence stimuli when combined with experiential ideation; participants were asked to imagine a positive or negative scenario as the piece played softly in the background [Eich et al., 2007]. This procedure was proposed

to employ both emotional musical exposure and self-determined emotional imagination in tandem. For example, one participant imagined walking their dog on a beach after their graduation, while exposed to Four Seasons: Spring III Allegro. Another participant imagined the funeral of a friend to the tune of Träumerei. Participants were directed to use their own imaginations in determining the content that they associated with each piece, while aware that it was to be either positively or negatively valenced.

The task requirements also provided the opportunity for a more structured conversation post-stimulus, and less room for subjective interpretation than video clips. The experimenter asked questions after each elicitation task to assess the effect of valence and engage the participant in using the avatar conversationally. For either valence stimulus, the faces of the respective avatars were displayed side-by-side in gallery view on the video call.

The individual stimuli duration was the same in this experiment as in the Pilot Experiment (90sec) and the total time spent in conversation with the avatar was 8 minutes in both experiments, only it was split into 4 segments for this version, as opposed to 2 conversation segments in the Pilot, due to the addition of two conditions in this experiment. Each emotion induction task was followed directly by a subjective questionnaire that participants had to fill in.

3.2.3 Measurements

Subjective data was gathered from a questionnaire at the end of each emotion induction task. Many of the question items represented in Table 3.1 were used for this second experiment's questionnaire.

The items aimed at measuring Enfacement, Social Presence and Emotion Elicitation. We included some questions attempting to measure the Proteus Effect and also questions for determining levels of fatigue post-task. They are summarized in Table 3.2, which also provides the abbreviations used in the results section.

Due to the quantitative increase in collection of data gathering via questionnaires for this experiment, no comment section was this time added to the forms. The questions were presented in different orderes dependent on the condition, to control for question order related biases.

The participants also completed the Ten-Item Personality Inventory (TIPI) [Gosling et al., 2003], prior to taking part in the experiment. This was done as an extra data gathering measure, so that we could investigate whether individual personality traits could have any observable impact on the effect scores.

Table 3.2: Main Experiment: Questions for the main experiment by variable name. Each question could be answered on a scale from 1 – "Strongly Disagree" to 5 – "Strongly Agree". Question with star were only asked in the *Avatar* condition.

Variable Name	**Question**
Agency1	*"The movements of my avatar's face were caused by my movements."*
Agency2	*"I felt as if my avatar's face were moving by itself."*
Agency3	*"I felt as if the movements of the virtual face were influencing my own movements."*
Ownership1	*"I felt as if the avatar's face was my face."*
Ownership2	*"It seemed like the virtual face belonged to me."*
Ownership3	*"It seemed as if I might have more than one face."*
Fatigue	*"I was fatigued after the task."*
Alert	*"I felt alert after the task."*
InteractivePresence1 *	*"The other avatar mimicked my movements."*
InteractivePresence2	*"The other person seemed stimulated by the conversation."*
Proteus1	*"I felt good natured as my avatar."*
Proteus2	*"I felt more friendly than usual during the experiment."*
Post *	*"When my avatar looked happy/Sad, the other avatar looked happy/Sad."*
SharedPresence1	*"I felt as if I was in the same room as the other person."*
SharedPresence2	*"The other person's emotions were clear to me."*
SocialPresence1	*"The conversation was like a face to face meeting."*
SocialPresence2 *	*"I mimicked the movements of the other avatar."*
Valence	*"When you described positive/negative events as your avatar, to what extent did you feel happy/sad?"*

Our Likert-scale data was reduced from 7 point to 5 point, to accommodate the larger volume of Likert data but also to ease cognitive load on participants, given they were being asked to report much more data than the previous experiments. The more defined path of the five point *Strongly Disagree-Strongly Agree* scale was chosen to standardise our data collection scores, given it is the same structure as the accepted TIPI assessment method [Gosling et al., 2003], and this method was also considered to be less taxing for participants to consider than the 7 point *Not at all-Extremely* scale used in the Pilot.

3.2.4 Hypotheses

In this experiment, we were first interested in investigating participants sense of enfacement towards an avatar in the context of dyadic video-conferences, and we expected participants to experience such senses towards their avatar while conversing with another person.

Because the scores of enfacement were relatively low for the avatar in the Pilot, we hypothesised that removing the avatar representation of the conversation partner (factor PartnerRepresentation) would encourage participants in engaging more with their avatar and therefore increase their sense of enfacement towards it. Because Proteus Effect is commonly associated with the sense of embodiment [Heide et al., 2013, Ratan et al., 2020], we also expected Proteus scores to be impacted by the representation of the conversation partner.

Furthermore, we aimed at exploring the impact of positive and negative valence (factor

Emotion) on participants senses of Social Presence, and other qualia related to these concepts. Work in this area has revealed that positive valence can be associated with higher levels of reported social presence [Bailenson et al., 2016]. We hypothesised that participants sense of Social Presence would be higher after the positive mood induction compared to the negative mood induction, and that having the partner representation removed would decrease this subjective feeling.

In addition, Valence question was introduced to ensure that the mood induction was effective, and we therefore hypothesised that when describing positive or negative events, participants would respectively feel happy or sad.

To summarise, our hypotheses for the experiment were the following:

- (H1) Participants will experience sense of enfacement towards their avatar during the dyadic video-conference.
- (H2) Participants' sense of enfacement will be higher in the Voice condition than in the Avatar condition.
- (H3) Proteus scores will be higher in the Voice condition than in the Avatar condition.
- (H4) Social Presence scores will be higher in the Positive condition than in the Negative condition.
- (H5) Social Presence scores will be lower in the Voice condition than in the Avatar condition.
- (H6) Valence scores will show evidence of mood induction efficiency.
- (H7) Positive mood induction will be more fully achieved with avatar representation of the partner compared to negative mood.

3.2.5 Results

Two-way ANOVA analyse were conducted, considering the within-group factor PartnerRepresentation (2 levels: *Avatar* and *Voice*) and Emotion (2 levels: *Negative*, *Positive*). The normality assumption was tested using Shapiro-Wilk test and when not verified, an Aligned Rank Transformation (ART) was applied on the data. Tukey's post-hoc tests ($\alpha = .05$) were conducted to check significance for pairwise comparisons. As for correlation analyses, Pearson's r (r) was used for parametric data and Spearman's r (r_s) was used for non-parametric data. In addition, post-hoc tests were corrected using Bonferroni correction.

To ease readibility, all significant main effects and interactions, as well as their effect sizes, are reported in Table 3.3 instead of directly inside the text.

Table 3.3: Main Experiment: Summary of main effects on subjective scores

Agency1: The movements of my avatar's face were caused by my movements.

Effect	Anova	η^2	Post-hoc
PartnerRepresentation	$F_{1,57} = 4.59,\ p = 0.04$*	0.18	Avatar > Voice p=0.04*
Emotion	$F_{1,57} = 1.17,\ p = 0.28$		
PartnerRepresentation:Emotion	$F_{1,57} = 0.74,\ p = 0.39$		

SharedPresence1: I felt as if I was in the same room as the other person.

Effect	Anova	η^2	Post-hoc
PartnerRepresentation	$F_{1,57} = 23.93,\ p < 0.001$***	0.42	Avatar > Voice p<0.0001***
Emotion	$F_{1,57} = 1.16,\ p = 0.28$		
PartnerRepresentation:Emotion	$F_{1,57} = 6.36,\ p = 0.01$*	0.32	Avatar Negative > Voice Negative p=0.005**

SharedPresence2: The other person's emotions were clear to me.

Effect	Anova	η^2	Post-hoc
PartnerRepresentation	$F_{1,19} = 5.34,\ p = 0.03$*	0.22	Avatar > Voice p=0.03
Emotion	$F_{1,19} = 0.43,\ p = 0.52$		
PartnerRepresentation:Emotion	$F_{1,19} = 15.00,\ p = 0.001$**	0.44	Avatar Negative > Voice Negative p=0.003**
			Voice Negative < Voice Positive p=0.03*

SocialPresence1: The conversation was like a face to face meeting.

Effect	Anova	η^2	Post-hoc
PartnerRepresentation	$F_{1,57} = 54.34,\ p < 0.001$***	0.56	Avatar > Voice p<0.0001***
Emotion	$F_{1,57} = 0.57,\ p = 0.45$		
PartnerRepresentation:Emotion	$F_{1,57} = 0.002,\ p = 0.96$		

Valence: When you described positive/negative events as your avatar, to what extent did you feel happy/sad?

Effect	Anova	η^2	Post-hoc
PartnerRepresentation	$F_{1,57} = 0.003,\ p = 0.96$		
Emotion	$F_{1,57} = 6.42,\ p = 0.01$*	0.47	Negative < Positive p=0.01*
PartnerRepresentation:Emotion	$F_{1,57} = 0.003,\ p = 0.96$		

Enfacement

Enfacement towards the avatar was assessed through items related to face-ownership and face-agency. Variables labelled Ownership 1-3 (Table 3.2) focus on face-ownership. There was no main effect observed within participants, showing that Ownership towards the avatar did not seem to be impacted either by the emotion induced or the representation of the conversation partner (see Figure 3.5).

A main effect of PartnerRepresentation was found on agency question Agency1 (see Table 3.3. Post-hoc tests highlighted that Agency1 scores were higher in *Avatar* condition compared to *Voice* condition, independently of the emotion induced *($p = 0.04$)* (see Figure 3.6, A).

This means that participants tended to feel more that the movements of their avatar's face were caused by their movements when they would have the possibility to see the partner's avatar compared to when they would not. Interestingly, it seems like the presence of the other's avatar has a positive influence on the agency of participants towards their avatar, which is the opposite of our (H2) assumption that enfacement would be stronger when not seeing the other's avatar. Overall, Agency1 being the only item influenced by PartnerRepresentation, and in the opposite way we had expected, our results do not support (H2).

Global Ownership and Agency scores were as well computed by combining scores of all the respective items. While the two-way ANOVA did not highlight any effect on those, their mean scores (Ownership: $M = 3.42, SD = 0.55$, Agency: $M = 3.16, SD = 0.63$) can be considered as relatively high. Combined, this evidence neither supports nor nullifies (H1).

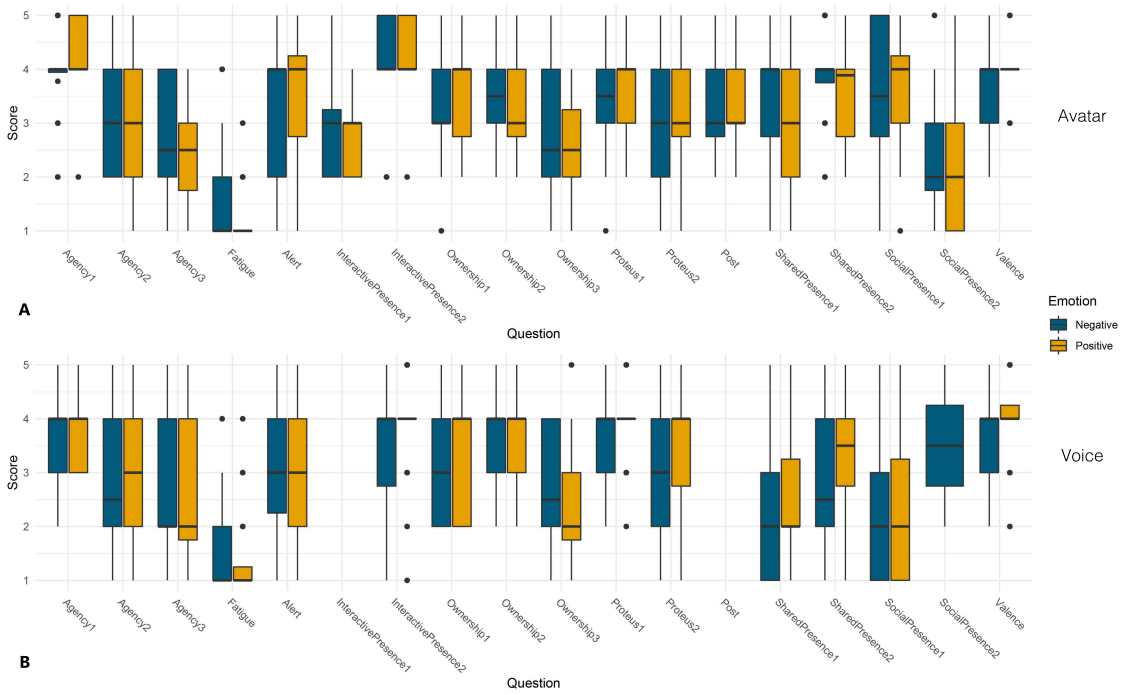


Figure 3.5: Main Experiment: Boxplots of all the dependent variables depending on factor Emotion for *Avatar* condition (A) and *Voice* condition (B).

Proteus Effect

Results on responses to item Proteus1 and Proteus2 in Table 3.3 showed an absence of main effect of our main conditions for improved self-concept alteration. Global Proteus1 ($M = 3.7, SD = 0.89$) and Proteus2 ($M = 3.24, SD = 1.16$) scores independently were relatively high, similar to the Enfacement scores and therefore coherent.

Social Presence

Social Presence was assessed through several items as "interactive presence", "shared presence", and "social presence" items, as presented in Table 3.2.

A main effect of PartnerRepresentation was found on the variable SocialPresence1 (see Table 3.4), confirmed by post-hoc tests *($p < 0.001$)* showing that participants felt more that the conversation was like a face to face meeting when they could see the avatar of the conversation partner, which was expected and goes in the direction of validating (H5) (see Figure 3.6, D).

Interestingly, the scores of SharedPresence1 and SharedPresence2 were also influenced by the PartnerRepresentation factor, but an interaction effect was found showing an impact of Emotion on these results (see Table 3.3). For SharedPresence1, post-hoc tests showed that the influence of PartnerRepresentation was only present when factor Emotion was *Negative ($p < 0.01$)* (see Figure 3.6, B). Negative valence therefore seems to influence how the representation of the partner will influence shared presence. In that case, participants felt more that they were in the same room as the other when the other had an avatar compared to when the other had only the voice, but only in *Negative* condition.

For SharedPresence2, post-hoc tests also showed that the influence of PartnerRepresentation was only present when factor Emotion was *Negative ($p < 0.01$)*, highlighting that participants found the emotions of the partner clearer with *Avatar* representation compared to *Voice* representation only when exchanging on negative content but not for positive valence (see Figure 3.6, C). When talking about positive content, the representation of the partner did not seem to have an impact on the recognition of the other's emotion.

This finding might be explained by the other post-hoc result, revealing an influence of Emotion when only considering the *Voice* condition *($p < 0.05$)*. When participants would not see the avatar of the partner, they would find the other person's emotions clearer when discussing positive content than when discussing negative content. This

might explain why PartnerRepresentation had only an impact in the negative condition, since in that case participants were finding it more difficult to recognize the partner's emotions. These results however do not support (H4).

Valence

Valence scores for all conditions included were relatively high ($M = 3.76, SD = 0.83$), showing that the mood induction was effective and validating H6.

In addition, a main affect of Emotion was found on Valence scores (see Table 3.3) and confirmed by post-hoc tests ($p = 0.01$) which showed that independently of PartnerRepresentation, participants felt more happy when talking about positive events than sad when talking about negative events (see Figure 3.6). This is consistent with the hypothesis that positive emotion induction was more naturally achieved with cartoon avatars.

Fatigue

While our analysis did not shown any effect of our conditions on participants fatigue, it seems like participants overall did not experience strong fatigue after the experiment ($M = 1.47, SD = 0.83$), however, they also did not report being particularly more alert after doing it ($M = 3.18, SD = 1.45$). Fatigue scores depending on Emotion are represented in boxplots for both *Avatar* and *Voice* conditions in Figure 3.5.

Personality Traits

It has been discussed in previous work that personality traits could influence users's sense of presence [Wallach et al., 2010] and sense of embodiment towards an avatar [McCreery et al., 2012, Dewez et al., 2019b] in virtual reality.

For this reason, we were interested in looking for possible correlations between the personality scores obtained with the TIPI questionnaire and subjective scores from the experiment. Table 3.4 outlines the observed results from the Spearman's test conducted, showing correlations of different strength depending on the variables.

The most striking significance recorded came from participants who scored highly on Agreeableness. There was a positive correlation observed between Agreeableness ratings and Ownership1 ($r_s = 0.30$, $p < 0.01$) and Ownership2 ($r_s = 0.26$, $p < 0.05$) questions. Moreover, Agency1 ratings were also positively correlated with Agreeableness scores ($r_s = 0.34$ $p < 0.01$). Although less strong, a similar positive correlation was found between Agreeableness and two social presence items SocialPresence1 ($r_s = 0.22$ $p < 0.05$) and SharedPresence1 ($r_s = 0.25$ $p < 0.05$).

Table 3.4: Main Experiment: Spearman correlations between personality scores and questionnaire items.

	Ownership1	Ownership2	Ownership3	Agency1	Agency2	Agency3	SocialPresence1	SharedPresen
Extraversion			0.26*					
Agreeableness	0.30**	0.26*		0.34**			0.22*	0.25*
Conscientiousness			-0.35*		-0.37**	-0.22*		
Emotional Stability								
Openness						-0.29*		

($* : p < 0.05$; $** : p < 0.01$)

There was also a positive correlation between Extraversion and item Ownership3 ($r_s = 0.26$ $p < 0.05$), showing that people that tend to be social and active are more likely to have a decreased sense of Ownership (Ownership3 being a control question: *"It seemed as if I might have more than one face"*). Curiously, results also showed that people with such personality traits had more chance to feel more alert after the task ($r_s = 0.23$ $p < 0.05$), but also more fatigued ($r_s = 0.25$ $p < 0.05$), which seems contradictory and suggest the potential interference of another factor yet not identified.

Conscientiousness scores were also significantly correlated with several Enfacement items, such as Ownership3 (r_s = -0.35 $p < 0.05$), Agency2 (r_s = -0.37 $p < 0.01$) and Agency3 (r_s = -0.22 $p < 0.05$). While the correlation is negative, these items being control questions, the results suggest that people that tend to be organized and reliable are more likely to experience a strong sense of enfacement towards the avatar. They are nevertheless less likely to feel more friendly than usual during the experiment, as shown by the negative correlation with Proteus2 (r_s = -0.32 $p < 0.01$).

Finally, Openness scores were negatively correlated with Agency3 (r_s = -0.29 $p < 0.05$), showing that participants who tend to be curious and creative were more likely to not feel as if the movements of the virtual face were influencing their own.

Gender

In order to explore the impact of gender on our result, two other independent variables were defined: Gender (whether participant was male or female) and GenderMatch (whether the gender of the participant and experimenter were the same or different). A mixed three way ANOVA was performed twice with Emotion and Avatar as within factors

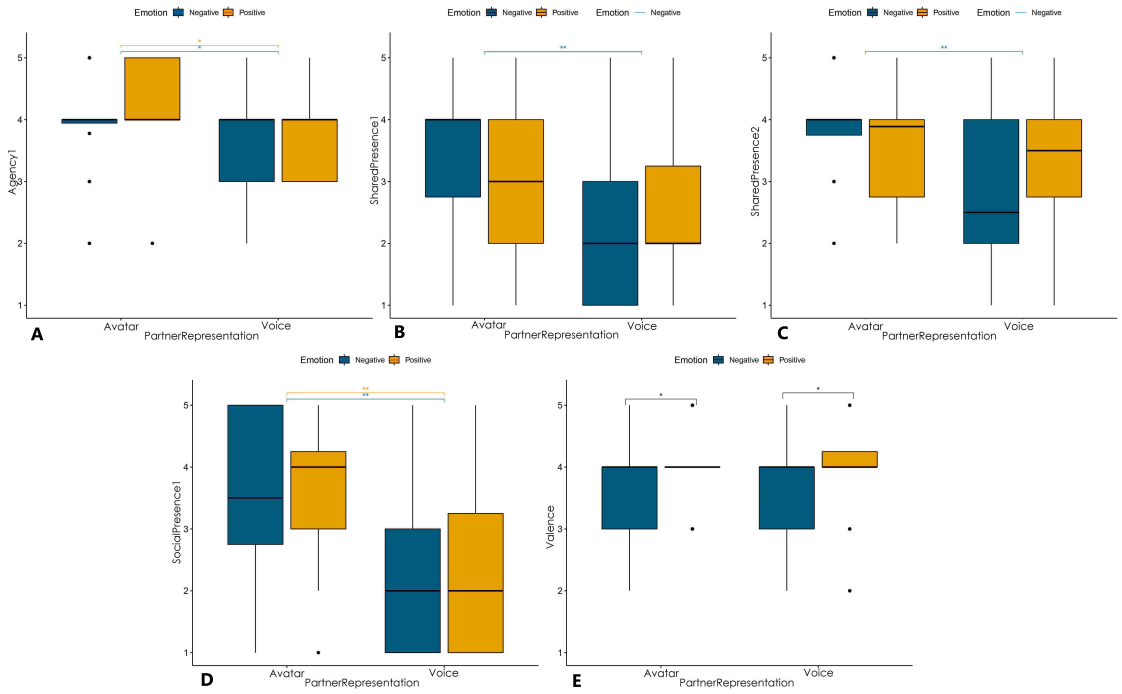


Figure 3.6: Main Experiment: Summary of significant effects of Emotion and PartnerRepresentation on Agency1 (A), SharedPresence1 (B), SharedPresence2 (C), SocialPresence1 (D) and Valence (E).

and either Gender (*Male* or *Female*) or Gendermatch (*SameGender* or *DifferentGender*) as a between factor.

A main effect of GenderMatch was found on SocialPresence1 variable ($F_{1,18} = 5.17$, $p = 0.04$) which was confirmed by post-hoc tests ($p = 0.04$) showing that SocialPresence1 scores were higher when participants did not have the same gender as the experimenter. A main effect of PartnerRepresentation was also found on SocialPresence1 ($F_{1,54} = 58.38$, $p < 0.001$) confirmed by post-hoc tests ($p < 0.01$).

Additionally, a main effect of Gender was found on Valence variable ($F_{1,18} = 5.15$, $p < 0.05$) confirmed by post-hoc tests ($p < 0.05$,) showing that Valence scores tended to be lower in the group of male participants rather that in the group of female participants.

3.2.6 Discussion for Experiment 1B

Compiled results emphasised complex consequences for our outlined Hypotheses H1-7. There were relatively high scores for our agency measures, which indicated participants felt more in control of their avatar during this experiment than in the Pilot. There was a measured interaction between PartnerRepresentation and Agency1, Table 3.2. This means that participants tended to feel more that the movements of their avatar's face

were caused by their movements when they would have the possibility to see the partner's avatar compared to when they would not, and is a direct contradiction to our second hypothesis (H2). Implications for this are discussed below. There were no other effects for enfacement between conditions, but scores were also not reported as consistently low (H1).

However, there was evidence from our Spearman's Correlation that participants who scored highly on Agreeableness tended to return higher results for face-ownership and agency. This is important for future work in this area, which might investigate further the interaction between individual personality traits and the reports of SoE.

Our results show that people who scored highly on Agreeableness were more likely to have a strong sense of ownership and agency towards their avatar, and therefore an enhanced enfacement feeling. They were also more likely to experience a strong sense of being in the same room as the other person and find resemblance with a face to face meeting, leading to a higher sense of social presence. Agreeableness has been implicated in previous studies of social presence, which have been comparably measured here [Dewez et al., 2019b].

Social presence was reported strongly in our results, suggesting that avatar-use can attain some of the social characteristics of in-person video-conferencing. The analysis revealed a measurable effect of social presence in the avatar use. While we confirmed that social presence was higher with avatar-avatar conversations (H5) we were surprised at observed results implicating negative valence interacting with the measured presence, more than positive valence (H4). It appeared that PartnerRepresentation was an influencial factor in negative conditions. This may be explained by volunteers reports of clearer emotion recognition for positive conditions.

We had to confirm the null hypothesis from our observations on Proteus effect measures (H3). The results did not display evidence for self-concept or behaviour change on our measurements, despite global Proteus results showing a consistently average scoring, rather than low scoring. There was an interesting relationship observed between TIPI and Proteus measurements, with high Conscientiousness scores negatively correlated with friendliness ratings post avatar condition.

With regards to valence conditions, the results for positive inductions were reported stronger than the negative ones. This is some confirmation of our hypothesis around mood induction (H6). Evidence also suggested that the cartoon avatars were more conducive to conditions of positive valence (H7).

Regarding the influence of gender, there was some evidence to suggest that non match-

ing pairs (e.g. Experimenter F, Participant M) were likely to consider the conversations more like face to face meetings that matching pairs. Moreover, valence scores tended to be lower in the group of male participants rather that in the group of female participants. While those results are difficult to interpret, they are not the first one suggesting differences between gender in subjective feelings of virtual experience, as previous works already highlighted the influence of such aspect on the sense of Ownership towards a virtual avatar [Dewez et al., 2019b] or on proxemics [Zibrek et al., 2020], which reinvigorates the need to further investigate that effect.

Furthermore, while there is not to our knowledge previous research exploring the impact of gender match between two individuals on virtual social presence, there seem to be an impact of gender matching in the process of learning in virtual platforms [Makransky et al., 2019], suggesting that further work exploring the impact of gender matching in social virtual contexts would be valuable for the community.

3.2.7 General Discussion for Experiments 1A and 1B

The experimental evidence described above shows a variation in observations on measured items. In general, it has been explicated that avatar-interaction studies for screen-based applications can provide actionable data around the quality of avatar experiences [Heide et al., 2013].

Social Presence

Observations from presence measurements seem to be the biggest contribution of the current research. Social presence results for both the Pilot and Main Experiment determinants implicate human use of an avatar can provide elements of the percepts underlying senses of presence.

The use of personalised avatars for remote collaboration has not been shown to detract from social presence in the main experiment, when we compared their use to non-video call conditions. The results shown here are consistent with previous work on presence and avatar personalisation [Waltemate et al., 2018].

Additionally, in cases where negative emotional content was induced, avatars were an improvement on voice condition. Emotion recognition in the Voice condition was comparably difficult for negative valence scenarios. Without the presence of the other avatar, it was more difficult for participants to recognise negative emotions rather than positive ones.

There was no decrease in presence measurements on the basis of emotional valence,

compared to Avatar conditions. Affected avatar studies such as ours have seen similar associations between positive affect and presence [Bailenson et al., 2016] which has important implications for future development of avatar software for video-conferencing requirements. Recording behavioural measures of social presence, such as mimicry of physical movements, through analysis of animation data and movement data correlations could be a useful point of future convergence for this kind of work.

Enfacing Avatars

While scores for face or body ownership observed in the Pilot were not particularly high, our re-formulated experiment saw a significant increase on measurements for agency, and slight increase for ownership. Tracking quality did not have an impact on agency in the Pilot.

The avatar condition in the main experimental design elicited higher scores for an Agency1, as participants felt more control over the movements of their avatar. This is a useful finding, as viewing one's own self avatar in a dyadic interaction with a voice agent could be implicated as having a potentially negative influence on feelings of agency compared to avatar-avatar conversations.

However, given that the participants could see both themselves and the other avatar (Avatar condition) or else their empty gallery (Voice condition) for the conversations that were measured, it is also likely that the view of mirror representation of the self-avatar was more strongly attended for Voice conditions. This mirror effect has been proposed to increase focus on animation details and hence decrease agency [Fribourg et al., 2018], which could explain our resutls.

The self-similarity of the virtual avatars for both experiences was not sufficient for producing high levels of enfacement, even though similarity and self-identity are considered key components to the experience of enfacement [Tajadura-Jiménez et al., 2012].

Furthermore, differences emerged for enfacement results between the Pilot and Main Experiment. Previous work has suggested enfacement measurements are not influenced by animation realism, which makes the absence of observed impact from our Quality measure in the Pilot experiment consistent with previous findings [Kokkinara and McDonnell, 2015].

Self-Concept Alteration

While there was some self-concept changes observed on friendliness for the Pilot, Proteus effects were not measured strongly in either the Pilot or the Main Experiment.

The results of the second study produced unconvincing measures of Proteus, but not a complete absence of ratings. It is interesting that enfacement measurements also yielded comparably underwhelming reports. The coextensive absence of effects in both ownership and self-concept results adds empirical credence to the notion that they must necessarily co-arise.

This implies some element of contingency between SoE and self-concept alteration, with embodiment implicated as a necessary precondition to perceived behaviour change which has been described by meta-analyses of Proteus effect and its interaction with other factors [Ratan et al., 2020]. The body is cognitively represented and volitionally experienced as an object mediated by perceptual content [Riva, 2018]. Changes in bodily percepts can result in changes to manifesting bodily behaviour. While the range of evidence varies between our experimental conditions, it has been explicated that behaviour change post self-identity change is a more complex phenomena in computer mediated avatar interactions than previously defined by research in Immersive VR [Heide et al., 2013].

Influence of Valence

In general the results displayed for valence measurements were informative, with efficacy observed in positive and negative mood inductions during avatar use. The valence results from our studies specifically emphasise an influence of positive affect and also indicate a relative weakness of affect from the negative stimuli in comparison. Personalised avatars have been recorded to influence with emotional responses [Waltemate et al., 2018], while positively affected avatars also have documented effects on dyadic conversations [Bailenson et al., 2016].

Our results suggest an interplay between these two measured phenomena, as the use of personalised cartoon avatars has indicated a strong positive valence induction here. Despite previous research claiming that negative inductions can be more easily achieved than positive ones [Ito et al., 1998], there is further useful work on remote mood induction procedures, that makes claims about the equitable efficacy for both positive and negative emotion elicitation via internet based methods [Ferrer et al., 2015]. Our results sit between these claims, and that of other avatar affect studies [Bailenson et al., 2016], as an example of mood induction efficacy that proposes to augment contemporary methods for remote valence induction.

Impact of Personality

Previous research on the influence of personality traits in immersive environments [McCreery et al., 2012] motivated our inclusion of the TIPI measurements for

the second iteration of these experiments. It proved to be informative. Our results on the correlation between Agreeableness and Agency, Ownership and Presence scores suggest some interaction occurred here, while Extraversion was correlated with Ownership Item 3, Table 3.2.

These results are promising, particularly because Agreeableness and Extraversion have been implicated in previous work on character realism in immersive settings [Zibrek et al., 2018], while Extraversion has also been correlated with immersive tendencies in such environments [Weibel et al., 2010]. Ownership has also previously been linked to the trait for Openness in a larger study [Dewez et al., 2019b]. Our results in tandem with the work outlined here indicate that the role of personality traits in virtual experiences can be a focus for researchers interested in quantifying subjective differences in effect quality for immersive experiences.

3.3 Experiment 1C: Subjective Perceptions of Human and Non-human Avatars

In the previous experiments, it was established that emotion communication with virtual avatars was achievable, had a tendency to be stronger for positively valenced scenarios, and in general could benefit from avatar representation, as opposed to just voice representation. Importantly, the voice by itself was also considered emotionally communicative without the presence of the avatar. Additionally, the results described several different effects of personality traits on user perceptions of embodiment. From this basis, a new line of inquiry could be drawn; could there be similar personality influences and reports of embodiment in virtual avatars that did not resemble real humans? Additionally, each of the participants in the previous experiment had varying degrees of prior experience with the use of avatars, and so a related question that emerged from our research was: could the experience of embodiment be swayed by exposure to the avatar concepts prior to embodiment?

For our next experiment, we considered users alone in embodying an avatar, without any vocal stimulus, at differing levels of generic visual realism. In other words, participants experienced two different avatar bodies; one human, and one non-human. Despite sharing the same bodily affordances (i.e. arms, legs, stationary sitting position), it was considered likely that the presentation of the avatars may still afford differing perceptions of embodiment, and that these perceptions were likely to be influenced by subjective individual traits. Additionally, given that the background research has informed us as to the nature of our construction of perceptions using learned concepts,

this work also employed a priming stimulus to test the effect of concept priming. The experimental format described below includes a form of supraliminal priming, where a priming stimulus was presented to subjects who were fully conscious of its presence [Bargh and Chartrand, 2000]. For both the control and primed conditions, the nature of the VR task remained only tacitly known by participants in advance. Priming was experienced prior to VE exposure, while the avatar experience remained in subsidiary as opposed to focal awareness [Polanyi, 1966, Longo and Bertenthal, 2009]. Our method draws upon the theoretical underpinnings of contextual priming [Cerda et al., 2021], which can be considered a form of concept priming [Nunez and Blake, 2003], with both methods having produced significant findings on the experience of presence in virtual reality, as described in the background section.

3.4 Experiment

In order to explore our research questions, an experiment was designed for users to embody two different avatars as a within subjects condition (see Figure 3.8) either with or without a prior priming stimulus as between subjects treatment.

3.4.1 Participants

Overall, 67 participants were recruited from two European universities. Two participants were unable to complete the experiment, and so analysis was conducted on 65 participants (38 Male 27 Female, mean age 23.5 years, 22 French native speakers, 43 English native speakers).The project received ethical approval (Application number: 16092022) prior to data collection. Participants received information and consent forms before the experiment began, and were asked to return consent prior to the lab appointment.

3.4.2 Protocol

After entering the experiment room, participants were asked to fill out the Ten Item Personality Inventory and Multidimensional Assessment of Interoceptive Awareness, Version 2., in order to gather trait and demographic data before the experiment began.

For the Concept Group, participants experienced the concept priming condition. Prior to the VE immersion, participants read a text article and viewed an accompanying image depicting a missing advertisement (human avatar) and an escaped advertisement (non-human avatar). The text and image were used to describe the visual details of the avatar before embodiment took place, with the gender of the avatar adjusted to match the participant. For the Control Group, participants did not experience any priming



Figure 3.7: Office scene for experiment with human avatar. The side mirror was included to show participants the movements of the avatar as mapped to their own movements.

conditions and proceeded directly to the VE.

The participants were administered the HTC Vive headset, with motion trackers placed in opposite corners on either side of the participant. The height of the desk and size of the avatar were adjusted in accordance with the respective height of participants.

Using the method of [Dewez et al., 2019a] for the VE and task, the VE was presented as a virtual office (See Figure3.7) in which participants were seated on a chair near a table, facing a screen and with a mirror to the side, where they could view their virtual avatar. Firstly, participants were invited to explore the motion of their avatar using the virtual mirror, where they could see their own upper body movements mapped to the movements of the avatar. When they were ready to begin, participants were asked to make movements with their hands following a series of shapes (a circle, an infinity sign.(See Figure 3.9) After moving their hands in the shapes displayed on the screen, participants were asked to fill out the Virtual Embodiment Questionnaire and the adapted State Mindfulness Scale. Directly after this questionnaire, participants were embodied in a second avatar, in the same virtual environment, and repeated the tasks. The avatar order was counterbalanced so that equal numbers of participants experienced different orders of avatar presentation, between ant and human, as well as equal numbers of primed or not primed conditions.



Figure 3.8: Three avatars embodied during the experiment. Participants chose either the female (right) or male (left) depending on their own gender, and were also embodied the the non-human avatar (center). The avatars were scaled to the size of each individual participant.

3.4.3 Measurements

The dependent variables chosen for this study are reflected in our choice of measurements. We were looking for changes in state body mindfulness between avatar conditions, and employed the *Body* component of the State Mindfulness Scale [Ruimi et al., 2022]. Our other between-conditions report measurement was the Virtual Embodiment Questionnaire (VEQ) [Roth and Latoschik, 2020], which provided data on SoE as perceived by participants. We expected SoE to vary between avatar conditions, with scores intended to be analysed together with priming conditions, as well as suggestibility and personality trait scores.

Outside of the VR task, participants also provided other personality and interoceptive awareness trait data, as cited above. These measurements provided a total of thirteen separate traits to analyse for effects on embodiment data. These traits included the aforementioned Big Five personality traits, as well as interoceptive awareness traits of Noticing, Worrying, Distracting, Attention Regulation, Emotional Awareness, Self Regulation, Body Listening and Trusting.



Figure 3.9: Example of shape trace task from the user's point of view.

3.5 Analysis

For both the SoE and SMS reports, items were analyzed both by averaging over the individual items for global scores, and by analyzing each item as an individual question.

We conducted a repeated measures ANOVA with within-subjects factor *avatar* (human, non-human) and between-subjects factors *priming* (yes, no) and *gender* (male, female) for every questionnaire item. Additionally, the same analysis was conducted for aggregate scores of Ownership, Agency and Change, in order to assess overall effects of priming and condition, as opposed to effects on individual items. There were no violations for Levene's equality of variance and thus, for post hoc tests, we conducted a Bonferroni interval adjustment, and report the modified pairwise comparisons.

Spearman's correlations were conducted to assess the relationship between individuals personality traits, as measured by the TIPI and MAIA, and their respective scores on the surveys for each of the human and non-human avatar conditions. These correlations were conducted using both global scores and individual item scores, as with the ANOVA.

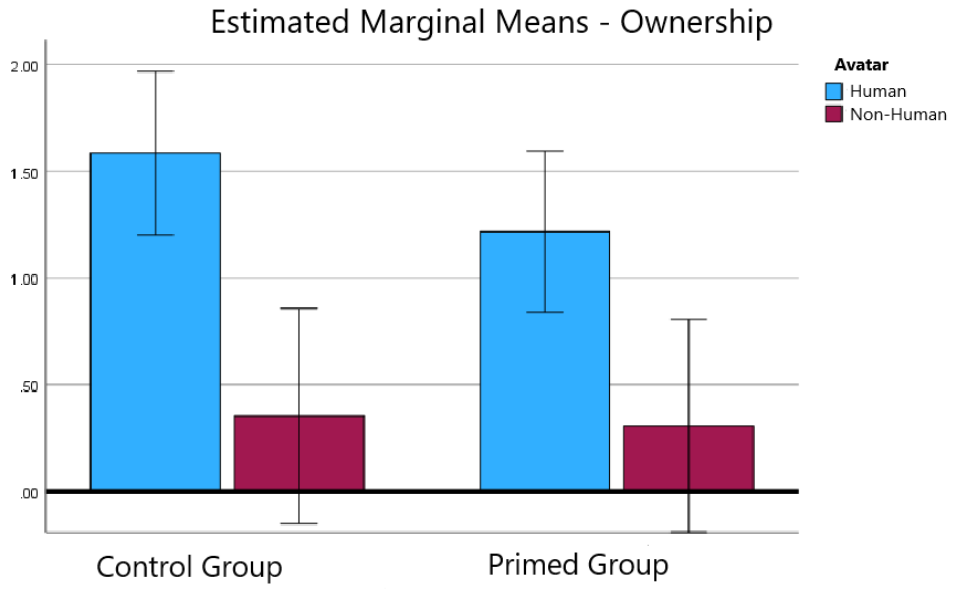


Figure 3.10: Estimated marginal means for Avatar, Priming and Ownership. Human avatar (blue) evoked significantly more ownership than non-human (red).

3.6 Results

Below we report significant results for aggregate and individual items, all other effects were non-significant.

3.6.1 Analysis of Variance

Ownership:

We first conducted an ANOVA on the aggregated items for Ownership and found a main effect of avatar ($F(1, 61) = 37.711, p < 0.001, \eta_p^2 = 0.382$) where higher levels of ownership were reported for the human than the non-human avatar ($p < 0.001$) 4.3.

Then, we conducted individual ANOVAs for each item to determine if any other effects were observed. We found a main effect of avatar for OW1 ($F(1, 61) = 11.913, p = 0.001, \eta_p^2 = 0.163$), OW2 ($F(1, 61) = 11.924, p = 0.001, \eta_p^2 = 0.923$), OW3 ($F(1, 61) = 50.183, p \approx 0, \eta_p^2 = 0.164$), and OW4 ($F(1, 61) = 13.093, p = 0.004, \eta_p^2 = 0.129$), where the human avatar was rated higher than the non-human on all items, OW1 ($p = 0.001$) OW2 ($p = 0.001$) OW3 ($p < 0.001$) and OW4 ($p = 0.004$). For OW4, we also found a main effect of Gender ($F(1, 61) = 40.243, p = 0.040 \eta_p^2 = 0.067$) where male participants felt like the virtual body belonged to them more than females ($p = 0.040$).

Agency:

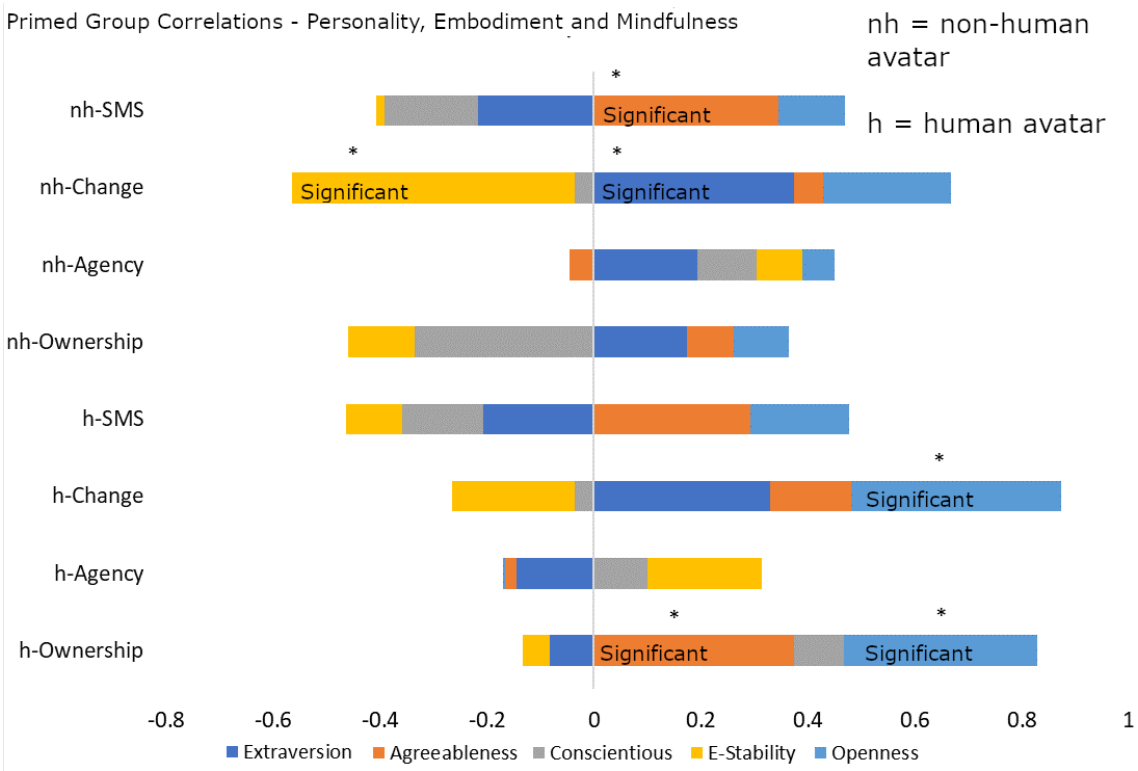


Figure 3.11: Plotted correlation coefficients for primed group personality traits with SoE and SMS results. Significant results labelled with a * for both human (h) and non-human (nh) avatar conditions.

For the aggregated items, we found a 3-way interaction between Avatar, Priming and Gender ($F(1, 61) = 4.925, p = 0.029), \eta_p^2 = 0.075$) where female participants experienced significantly more agency for the human avatar in the control condition ($p = 0.030$), while there was no significant difference in agency for the priming condition. In comparison, there was increased agency in the non-human condition for females in the primed group.

For the individual item analysis, we found a main effect of avatar for AG1 ($F(1, 61) = 4.259, p = 0.043, \eta_p^2 = 0.065$) where the movements of the virtual body felt like they were the participants movements more for the human avatar than the non-human avatar ($p = 0.043$).

For AG2, we found a 3-way interaction between avatar, priming, and gender ($F(1, 61) = 5.263, p = 0.025, \eta_p^2 = 0.079$) which showed a greater sense of control of the non-human body for Males in the Control condition compared to the Primed condition ($p = 0.024$).

For AG3, we found a 2-way interaction between avatar and gender ($F(1, 61) = 5.730, p = 0.02, \eta_p^2 = 0.086$) where perception of causing avatar movement was higher for female participants in the non-human avatar than in the human avatar ($p = 0.042$), while there was no effect for male participants.

There was also a main effect of Avatar for AG4 ($F(1, 61) = 4.909, p < 0.001, \eta_p^2 = 0.074$) where participants felt more in sync with the movements of the non-human avatar compared to the human avatar ($p < 0.001$). Additionally, there was an interaction effect between Avatar and Gender ($F(1, 61) = 4.909, p = 0.018, \eta_p^2 = 0.074$) where female participants felt more in sync with the movements of the human avatar compared to males ($p = 0.029$).

Change:

For the aggregated items, we found no main effects or interactions.

For the individual item analysis, we found a main effect of avatar for CH4 ($F(1, 61) = 4.280, p = 0.043, \eta_p^2 = 0.066$) where participants felt that the width of their own body had changed more for the non-human avatar. No other main effect of avatar.

There was a main effect of gender for CH4 ($F(1, 61) = 7.438, p = 0.008, \eta_p^2 = 0.109$) where female participants felt that the width of their own bodies had changed more than males ($p = 0.043$).

State Mindfulness Scale: There were no main effects for overall SMS scores. There was a main effect of avatar for SMS8 ($F(1, 61) = 5.507, p = 0.022, \eta_p^2 = 0.083$) where participants felt that they more clearly physically felt what was happening in their body

after embodying the human avatar ($p = 0.022$). There was also a main effect of Avatar for SMS18 ($F(1, 2.466) = 5.349, p = 0.024, \eta_p^2 = 0.081$) where participants felt more in touch with their body after using the human avatar ($p = 0.024$). There was also an Interaction between Avatar, Priming and Gender for SMS18 ($F(1, 1.874) = 4.064, p = 0.048, \eta_p^2 = 0.062$). Male participants in the primed group felt more in touch with their body after the non-human avatar ($p = 0.003$).

3.6.2 Correlations

Personality Inventory:

Before analysing individual items across both priming and non-priming groups, we first report correlations between aggregate items for sense of embodiment and state body awareness metrics with the items measured by the TIPI and MAIA.

For the primed group (N = 33), there were significant positive correlations for Extroversion and Change for the non-human condition ($r = .374, p = 0.032$), Agreeableness and Ownership for the human avatar condition ($r = .375, p = 0.032$) as well as body mindfulness for the non-human condition ($r = .346, p = 0.049$), and Openness with both Ownership ($r = .362, p = 0.038$) and Change ($r = .393, p = 0.024$) for the human condition. There was also a significant negative correlation between Emotional Stability and Change for the non-human condition ($r = -.529, p = 0.002$) (3.11.

Surprisingly, there were no significant correlations recorded for the control group (N = 32).

When analysing each individual item, significant positive correlations were observed between trait Extroversion and SoE Item CH3 for the human avatar condition ($r = .308, p = 0.013$) and CH3 in the non-human avatar condition ($r = .336, p = 0.006$). There was also a significant negative correlation for SMS14 for the non-human avatar condition ($r = -.265, p = 0.033$).

Trait Agreeableness was positive correlated with scores for item OW2 ($r = .266, p = 0.032$), SMS8 ($r = .255, p = 0.040$) and SMS14 ($r = .335, p = 0.006$) in the human avatar condition, as well as SMS14 ($r = .412, p > 0.001$), and SMS18 ($r = .258, p = 0.038$), in the non-human condition.

For the measured trait of Conscientiousness, there was a positive correlation recorded for CH3 ($r = .252, p = 0.043$) in the human condition.

Emotional Stability was negatively correlated with SoE item CH4 in the non-human avatar condition ($r = -.282, p = 0.023$).

Trait Openness was correlated with the most items, including positive correlations in the human avatar condition for items OW1 ($r = .254, p = 0.041$), OW2 ($r = .275, p = 0.027$), OW3 ($r = .273, p = 0.028$) and CH3 ($r = .414, p > 0.001$). Additionally there was a positive correlation for SMS9 in both human ($r = .276, p = 0.026$) and non-human ($r = .350, p = 0.004$) conditions.

Interoceptive Awareness:

In the primed group, there were significant positive correlations for MAIA items Worrying and Agency for the human avatar condition ($r = .449, p = 0.009$) Emotional Awareness and Change for the non-human condition ($r = .406, p = 0.020$) Self regulation with both Agency ($r = .494, p = 0.004$) and Change ($r = .353, p = 0.044$) for the non-human condition, as well as trait body listening with Agency ($r = .364, p = 0.037$), Change ($r = .404, p = 0.020$) and body mindfulness ($r = .392, p = 0.024$) for the non-human condition. There was also a significant negative correlation between Worrying and Change for the non-human condition ($r = -.374, p = 0.032$).

For the control group, there were significant positive correlations observed for trait Worrying with both Ownership ($r = .390, p = 0.028$) and Agency ($r = .497, p = 0.004$) in the human avatar condition, trait Self-Regulation with Ownership ($r = .365, p = 0.040$) and Body Mindfulness ($r = .446, p = 0.011$) for the human condition, as well as Body Mindfulness for the non-human condition ($r = .433, p = 0.013$).

We also analyzed each individual item separately. In the human avatar condition, there was a significant positive correlation between trait Noticing and body ownership item OW2 ($r = .332, p = 0.007$). Other results were observed for AG3 ($r = .265, p = 0.033$) and AG4 ($r = .305, p = 0.014$).

Trait Worrying had several significant positive effects on embodiment reports in the human avatar condition, namely OW3 ($r = .296, p = 0.017$), AG2 ($r = .390, p = 0.001$), AG3 ($r = .391, p = 0.001$), AG4 ($r = .358, p = 0.003$) as well as negative correlations for CH3 ($r = -.291, p = 0.019$) and state mindfulness item SMS13 ($r = -.286, p = 0.007$). For the non-human condition, there were negative correlations recorded for OW3 ($r = -.294, p = 0.018$), CH2 ($r = -.353, p = 0.004$), CH3 ($r = -.347, p = 0.005$), CH4, ($r = -.300, p = 0.015$) and SMS13 ($r = -.258, p = 0.038$).

Attention Regulation had two significant positive correlations for the human avatar condition, namely OW2 ($r = .339, p = 0.006$) and CH3 ($r = .277, p = 0.026$).

Emotion Awareness was positively correlated with OW4 ($r = .282, p = 0.023$) and CH1 ($r = .297, p = 0.016$) for the human avatar condition, as well as a negative correlation

with SMS13 ($r = -.245, p = 0.049$). For the non-human condition, there were positive correlations with OW2 ($r = .252, p = 0.043$), OW4 ($r = .251, p = 0.044$), CH1 ($r = .277, p = 0.025$), CH2 ($r = .281, p = 0.023$), CH4 ($r = .254, p = 0.041$) and SMS9 ($r = .353, p = 0.004$).

Trait Self Regulation was positively correlated with OW1 ($r = .282, p = 0.023$), OW2 ($r = .345, p = 0.005$), OW4 ($r = .290, p = 0.019$) and SMS9 ($r = .389, p = 0.001$) in the human avatar condition, as well as CH2 ($r = .287, p = 0.021$) SMS9 ($r = .325, p = 0.008$) and SMS18 ($r = .281, p = 0.023$) in the non human condition.

Two positive correlations for trait Body Listening were observed for the non-human condition, in OW1 ($r = .267, p = 0.032$) and OW2 ($r = .272, p = 0.028$).

3.7 Discussion for Experiment 1C

There were two main objectives of this experiment. First, we aimed to deploy an adapted priming method to explore the external factors which can affect virtual embodiment scenarios. The second focus was to assess whether subjectivity in personal traits might also be influential on reports of perceived embodiment. Primarily, the Sense of Embodiment reports indicate a preference for human avatars in terms of ownership, as discussed by Kim et al [Kim et al., 2020]. Moreover, each of our hypotheses were at least partially supported by our results.

RQ1: Does concept priming prior to VR exposure improve SoE?

H1: Given previous reports of concept priming methods improving the sense of presence in VR, we expected higher reports of SoE from the concept priming group compared to the control group.

Some evidence exists in our correlation data to support H1, although our ANOVA results appear to support the null hypothesis. However, our findings suggest that a heightened sense of embodiment, as related to concept priming, is heavily dependent on individual factors.

The most significant finding in terms of analysis of variance was that priming had a negative effect on the sense of agency for female participants in the human avatar. This result constitutes evidence that our method of concept priming brought the avatar into the focal awareness of female participants. This awareness appears to be culpable for a reduction in agency, which may be due to the avatar concept highlighting the interruption

of interoceptive signals that indicate agency. In other words, for the human avatar condition, it is possible that the awareness of the avatar-as-avatar as opposed to avatar-as-afforded body meant that female participants felt less control of the avatar and its movements. Conversely, however, this result appears to suggest that agency for the non-human avatar was higher in the primed condition for female participants, which implies a role for humanness in dictating perceptual responses to concept priming for afforded bodies. Interestingly, similar gender specific results on SoE for female participants were found by Dewez et al [Dewez et al., 2019a].

The results from our correlation analyses paint a different picture in terms of the subjectivity of priming effects. There is evidence to suggest that priming influenced embodiment and body mindfulness reports for participants who were higher on traits Extroversion, Openness and Agreeableness, for both human and non-human avatars. Additionally, for MAIA traits there were significant correlations between Emotional Awareness, Self Regulation and Body Listening with perceived body change in the non-human avatar, that were present for the primed group and not present for the control group. The implication in these findings is that these particular traits are potentially influential on our responses to non-human embodiment. This contingency may be due to the subjectivity inherent to bodily perceptions in general, which are then mapped to our experiences of virtual bodies.

For the Five Factor Model traits, there is an intuitive interpretation to be found in the idea that positive prosocial behaviours, such as the externality of persona, the ability to welcome novel experiences, and the tendency to agree to predefined patterns, are influenced by simple concept priming in a way that enhances the embodiment experience. Similarly, the findings that three awareness focused traits in the MAIA enhance the perception of body change may highlight an experience of awareness that allows for the flexibility of interoceptive signals when it comes to afforded embodiment. Moreover, it is plausible to suggest that certain external or awareness focused traits may predispose people to have a heightened sensitivity to perception and cognition of concepts. These interpretations must be interpreted as preliminary rather than elementary, as confirmatory studies with larger samples would be required to validate the findings above. However, these findings are novel and potentially impactful on the development of VR applications, and as such may motivate further research on the subjective experiences of both human and non-human avatar embodiment.

Previous works which have employed the use of text and picture based articles to prime participants for VE immersion [Nunez and Blake, 2003, Cerda et al., 2021] have found significant effects on the sense of presence. While embodiment and immersion can

be understood as similar phenomena, the important difference arises with the sense of self that is naturally attuned to bodily states. This can be seen in our analysis of variance findings, as the female participants experienced less agency when primed for human avatar embodiment. While concept priming for virtual environments appears to be effective for increasing the sense of presence, the same method does not necessarily achieve the same effects for virtual embodiment, although there is reason in our results to explore further as to whether this could be the case for non-human embodiment. Moreover, the above cited previous works in this domain do not account for the individual differences which may influence sensitivity to primed concepts, as suggested by the correlation analyses. In general, our results on priming suggest an effect similar to that of Longo et al [Longo and Bertenthal, 2009]. Despite a difference in priming method and overall focus, there are similarities to be drawn from the easy acquisition of concepts which we use to predict affordances in embodied, social or object oriented interactions. Finally, an additional contribution may be found in the subjectivity of the observed effects; people acquire bodily affordances in different ways, depending on their respective relationships to their bodies and to the world.

RQ2: Will participants experience the same level of body mindfulness in both human and non-human avatars?

H2: We expected body mindfulness to be more difficult to maintain for the non-human avatar.

From our results, there appears to be support for H2. There were main effects for two SMS body items when analysed in isolation, which suggests that non-human avatar embodiment could hinder body mindfulness. Our interpretation of these results is that there is a more natural perceptual mapping from human bodies to human avatars, which requires more cognitive resources to achieve when the avatar is non-human. Hence, body mindfulness is more difficult to maintain.

Other changes in body mindfulness appear to be dependent on personality traits. From the Five Factor Model, there was a significant correlation between trait Agreeableness and state body mindfulness in both human and non-human avatar conditions, specifically in the primed group. As mentioned previously, it makes intuitive sense that trait agreeableness may aid in the maintenance of body mindfulness regardless of changes to exteroceptive cues and bodily affordances.

Additional results may be interpreted from the findings of MAIA traits. There is evidence that both primed and control groups were more likely to experience greater body mindfulness dependent on degrees of awareness and control, as Self-Regulation corre-

lated with body mindfulness for the control group, in both avatar conditions, while trait Body Listening correlated with body mindfulness for the non-human avatar in the primed group. It seems likely that self-regulation would be useful for maintaining coherence of bodily cues across both virtual and analog affordances, which in turn could make visual differentiation between human and non-human avatar bodies to be arbitrary in determining body mindfulness. Likewise, body listening may be useful for maintaining body awareness while embodying a non-human avatar.

While previous studies have identified relationships between SMS body mindfulness measurements and reports of ownership and agency [Döllinger et al., 2022], our study was among the first to explore reported state mindfulness in conjunction with challenged embodiment and personality traits. As such there are novel findings in our data which expand on the conclusions of Döllinger et al [Döllinger et al., 2023] by assessing state body mindfulness in a virtual task scenario with non-personalised avatars.

RQ3: Which personality traits influence the sense of embodiment?

H3: We predicted that trait Agreeableness would positively correlate with Ownership.

While there exists enough evidence to confirm H3, there were additional findings, as well as significant effects of personality traits in both the human and non-human avatar conditions.

Our main prediction on Agreeableness was based on the prior work of experiment 1B, and this finding was replicated in this work for the human avatar.

For the human avatar condition, there was also a significant effect of trait Openness on body ownership reports. This finding is both novel and intuitive, while adding to previous work which has correlated Openness with improved task performance in VR [Katifori et al., 2022]. This result suggests that trait Openness may improve the reaction to the affordances of virtual bodies. Adopting a virtual body may interrupt regular interoceptive signals that link bodily sense perceptions to a coupled environment, as normal bodily and environmental perception is forced to shift in the case of VE immersion. Trait Openness may aid in the perception of this transition, as it is likely that novel bodily affordances are treated more readily as regular bodies by subjects who are more open to the experience.

For the non-human condition, there was a significant effect of Extroversion on the perception of body change, predominantly to do with perceived change in size. As such, trait Extroversion may be implicated in improving the sense of body change in non-human avatars, although significant confirmatory research would be required to verify

that link.

With our correlation data, it is notable that the three traits measured to positively correlate with SoE and body mindfulness in the primed group were the three externally motivated personality traits of Extraversion, Openness and Agreeableness. Conversely, the negative correlation with perceived body change was one of the internally motivated traits, namely Emotional Stability. There is an interesting problem to be explored in this finding, in that our relationship to external objects, as defined by our respective personas, may positively influence our relationships to those objects and our access to their affordances.

Of primary interest in analysing MAIA traits against embodiment scores was to expand the range of known influences of individual subjectivity on the sense of embodiment. There were two novel results that were specific to the non-human avatar, as well as novel significant effects for the human avatar. In the non-human condition, traits Emotional Awareness and Body Listening were positively correlated with multiple SoE measurements, amounting to influences on agency and ownership respectively. This result expands the range of measured trait influence in non-human avatar embodiment. A possible explanation might be that the ability to pay attentive control to interoceptive body signals allows for an easier transition to driving external avatar bodies. Further studies which aim to assess how the subjectivity of interoceptive signals interferes with the sense of embodiment would be required to fully explicate this finding.

In these findings, there is sufficient evidence to suggest that there are significant trait influences to be found in avatar embodiment scenarios. On the back of prior work, these findings suggest that trait assessments are important to contextualise future studies on avatar embodiment. Of course, in order to confirm this notion analyses will need to be conducted on much larger samples, with greater variation in embodiment conditions.

One notable limitation of this study stems from the old adage about correlation data. Future studies which aim to verify our findings could benefit if they were designed to analyze data using other kinds of statistical models.

Another limitation may be drawn from our definition of Sense of Embodiment, as defined by [Roth and Latoschik, 2020], whereby we included items for body change instead of questions of self-location. Comparisons with other SoE questionnaires and measurement methodologies would be a useful development on this issue.

A significant limitation of this study would be our distinction between human and non-human avatars. While the two conditions differed significantly in terms of visually salient information, the functional affordances that were offered by the kinematic mapping were

the same for both human and non-human embodiment. Future research that investigates whether changing some of those motoric affordances, in a larger or smaller avatar, or else in a quadrupedal avatar, would be of benefit in addressing this limitation. Further studies may also seek to address the general interaction challenges inherent to non-human avatar embodiment, specifically how these motoric affordances may be perceived when a non-human embodiment is combined with relevant haptic and tactile elements in a VE [Cheymol et al., 2023]. Finally, our findings on concept priming may be relevant for any works that seek to aid the transition between real and avatar bodies, or indeed between different avatar bodies, as can be achieved through both active and passive transitions [Otono et al., 2023].

3.8 Conclusions

Overall there were numerous contributions made by the above works to the study of human-avatar interactions. The main contributions showed a preference for human as opposed to non-human avatars, as well as positive valence preferences and the subjective effects of personality traits on senses of body ownership and social presence.

These findings are unique and specific to the experimental designs at hand. Experiment 1B described above appeared to show little by way of difference between pure voice only emotional interactions and human-controlled avatars combined with voices, in terms of perception and reaction to emotionally valenced content. In experiment 1C, the subjectivity of avatar embodiment was highlighted in our trait results, which was considered a useful line of inquiry for our next sets of experiments. While visual appearance and more realistic portrayals of human characters appeared to dictate the relationship to the avatars, in that perceived embodiment was significantly greater for humanlike avatars, there was a minor effect of concept priming observed for our female participants, specific to a reduced sense of agency. The implication of this result is considered to be influential on the relationship between a user's body and an avatar body.

These findings, taken together, encouraged us to take a more holisitic view on the perceptions of these characters; the vocal and visual affordances provided in the avatar interactions were observed to be efficacious on the perceptions of those characters and the engagement with emotionally valenced content, as well as with the embodiment of the characters and the attentiveness to primed concepts. Hence, our focus shifted from the embodiment of these characters towards an embodied view of emotion perception in general.

What follows in the next sections are experiments which were designed to address several

specific lines of inquiry, as well as the following broad questions; What about the vocal and visual characteristics afforded emotion perception? And how might this aspect of perception change under differing conditions of realism? These questions are addressed by the following sections.

4 Experiments 2A-C: The Effect of Unimodal Realism Variations on the Perception of Virtual Humans

Our avatar interaction experiments provided numerous contributions to the field of human-computer interaction, with our main findings indicating a preference for positive valence and human-like visual appearances, with evidence to suggest that a human voice was also suitable for emotional scenarios. From this point, our questions deepened; it was established that these factors influenced virtual character perceptions, but we did not yet know how variations in these factors might heighten or dampen sensitivity to emotional expressions. The following experiments were designed to further explore this line of questioning. Experiment 2A was conducted in order to assess how visual realism variations could influence perceptions of emotion intensity and the Uncanny Valley when virtual faces were combined with a real human voice. Experiment 2B took the opposite approach: three different voices were applied to one realistic character to assess perceptions of the Uncanny Valley, empathy and vocal consistency. Finally, Experiment 2C was designed to assess whether visual realism variations in emotionally expressive characters could provoke more or less empathy when they were not paired with vocal expressions.

4.1 Experiment 2A: Visual Realism Variations with Real Voices

Our first experiment with independent virtual humans aimed to test whether substantially different apperance realism conditions could produce differing perceptions of virtual characters, provided that vocal affordances for emotion perception remained consistent.



Figure 4.1: Characters 1–6. Range of male (left) and female (right) virtual humans used in experiment, shown in the high fidelity condition.

This evaluative study was proposed to identify whether perceptions of the Uncanny Valley could be observed to differ between conditions of visual realism for characters with human voices.

The measurements we employed for quantification of the uncanny valley were derived from a revised alternative [Ho and MacDorman, 2010a] to the Godspeed indices [Bartneck et al., 2007] which can be described as central to the development of experimental techniques for measuring the uncanny valley. The three indices defined in this literature that we used for our measurements were *Human-Likeness, Eeriness* and *Attractiveness*. The valence scales were derived from work similar to ours [Mäkäräinen et al., 2014] which aimed to quantify emotional intensity in virtual characters.

The method outlined here was developed to examine perceptions of highly realistic virtual humans. The experimental format was constructed with 2 (Appearance fidelity) x 2 (Emotional Valence) conditions, with 3 character variations and 2 character gender conditions for a within subject design. Our main hypothesis was based on appearance fidelity, as we considered it likely that the maximal realism of the highest level of detail to be sufficient for traversing the uncanny valley in regards to virtual humans. We expected levels of uncanny ratings to differ between valence conditions. The experimental paradigm developed here was geared towards exploring potential evidence for and against these hypotheses.

4.1.1 Stimuli and Procedure

The stimuli were created in Unreal Engine 4.26 (UE4.26)[1]. The avatars used in the stimuli are MetaHumans created using the Unreal Engine online application MetaHuman Creator[2]. MetaHumans are currently the state-of-the-art for realistic digital humans. Each MetaHuman is available in 8 levels of detail (LODs), with LOD0 having the highest

[1]https://www.unrealengine.com
[2]https://metahuman.unrealengine.com

level of fidelity and LOD7 having the lowest. The differences in fidelity between LODs arise from differences in geometry complexity (for example, 24000 facial vertices at LOD0 versus 1300 at LOD4), texture complexity, and the use of strand-based hair (LOD0) versus card-based hair (LOD4). Each MetaHuman has a full facial and body rig for animation, with the complexity of these rigs also varying across the different LODs. For example, the LOD0 facial rig consists of 713 joints and 669 blendshapes, whereas the LOD4 facial rig has only 84 joints and no blendshapes. An identical rigging setup across MetaHumans means that a given animation can be used on multiple MetaHumans.

For each gender, we used three MetaHumans chosen on three levels of character material variation, representing three different ethnicities. This decision was motivated by previous studies investigating the uncanny valley [Macdorman et al., 2008] which have ascribed low ratings of eeriness to photorealistic virtual faces but have described the limitation of using only a single style of possible human variation. The character material and gender variation were also included to account for any potential biases in the participant sample. We used these six MetaHumans with LOD0 for our high fidelity condition (Figure 4.1 (top)) and LOD4 for our low fidelity condition. For our purposes, LOD 4 was considered an appropriate choice for low fidelity, representing the standard of previously gold standard renderings of human likenesses, in terms of both polygon count and appearance fidelity. LOD4 (Figure 4.1 (bottom)) is also considerably less humanlike than LOD0.

For each emotional valence, two trained actors' (one female, one male) facial performances were captured via a desk-mounted video camera placed in front of their faces. The valence expressions were derived from literature outlining particular phrases that may be used to represent positive or negative valence emotions [Ben-David et al., 2011]. Each recorded performance was then tracked using the Faceware Studio facial performance tracking software[3] and streamed into UE4.26 using the Live Client Plugin for UE4.26[4]. The facial performance was then mapped to the MetaHuman facial animation rig and the resulting animation recorded using UE4.26's Sequence Recorder feature. The commonality of the facial rig across MetaHumans meant that for each emotional valence a single recorded animation could be used across all female characters and similarly for the male characters.

The different components of the stimuli (MetaHuman, animation, lighting, virtual camera, and audio) were pieced together using UE4.26's Sequencer feature. The same virtual camera setup and standard three-point lighting setup was used for all stimuli.

[3] http://support.facewaretech.com/studio
[4] https://glassboxtech.com/products/live-client

For each gender and emotional valence, the same audio track and animation was used, meaning the only difference between stimuli was the actual MetaHuman itself. Several post-processing effects such as ray-tracing and anti-aliasing were used on the final stimuli which were rendered using the UE4.26 Movie Render Pipeline. Each stimuli was rendered at 30 FPS and at a resolution of 2048×1536 and downsampled to 1024×768. A Nvidia GeForce RTX™ 3090 GPU was used for the rendering.

20 participants (10 male, 10 female) were recruited from a combination of university mailing lists and Prolific participant recruitment[5]. These recruitment methods have been validated as appropriate and useful for perceptual research [Jonell et al., 2020]. In consideration of the complexity of ethical standards around paid participant research [Ripley, 2006] we paid the national minimum wage to Prolific recruits, adjusted to the duration of the experiment. The participants were provided with a link to complete the experiment in their web browsers, using Qualtrics survey software[6]. The duration of time taken to complete the experiment was considered as a means of rejection; the maximum accepted timespan was 48 minutes. This was to ensure insofar as possible that participants engaged in the experiment consistently throughout.

4.1.2 Results

We initially conducted a 3-way Analysis of Variance to assess the influence of within group factors *Fidelity* (High/Low), *Character* (1–6) and *Clip* (Positive/Negative) on reported levels of uncanniness and emotion intensity. However, there was no evidence in the data to suggest Character had any statistically significant effect on uncanniness and emotional intensity and was thus removed from the analysis to improve interpretability. A Shapiro-Wilk test for normality was performed on the data. The normality assumption was violated ($p>.05$), so we applied an Aligned Rank Transformation to the data as a means to enable our analysis of variance. For the scales of emotion intensity, we averaged results over character and ran 2 way analysis of variance on ratings of emotion intensity for factors of *Fidelity* and *Clip* as there we were not predicting individual differences from characters on emotion intensity.

Tukey's post-hoc tests ($\alpha = .05$) were conducted to check significance for pairwise comparisons.

In addition, post-hoc tests were corrected using Bonferroni correction.

- *Human-Likeness:* We observed a main effect of Fidelity on human-likeness

[5]https://www.prolific.co/
[6]https://www.qualtrics.com/uk/

($F_{1,437} = 117.96$, $p < .001$), indicating that participants considered high fidelity characters to look more human ($M = 4.12, SD = 1.82$) than the lower fidelity characters ($M = 2.87, SD = 1.58$) ($p < .0001$).

- *Eeriness:* A main effect of Fidelity was observed on questionnaire responses for Eeriness ($F_{1,437} = 30.97$, $p < .001$), with low fidelity characters rated as more eerie ($M = 4.05, SD = 1.66$) across the board than high fidelity characters ($M = 3.39, SD = 1.65$) ($p < .0001$).

- *Attractiveness:* A main effect of Fidelity on Attractiveness was established ($F_{1,457} = 120.6$, $p < .001$). Participants rated Attractiveness to be higher in high fidelity characters ($M = 4.74, SD = 1.41$) than low ($M = 3.6, SD = 1.41, p < .0001$), ($p =< .0001$). A main effect of Character on Attractiveness was found ($F_{1,437} = 7.98$, $p < .001$). As expected, some characters were rated more attractive than others, with a tendency of character 1 and 2 (Figure 4.1) to be less attractive than others and character 4 to be more attractive ($p < .0001$).

- *Happiness Intensity:* A main effect of Fidelity was found ($F_{1,19} = 14.45$, $p = .001$) on happiness intensity ratings, where ratings were higher for high fidelity characters than low ($p = 0.0012$). A main effect of Clip was found ($F_{1,19} = 137.96$, $p < .001$), where happiness intensity was rated higher in the positive than the negative clip, as expected ($p < .0001$). An interaction between Clip and Fidelity was found ($F_{1,19} = 5.81$, $p = .03$), where post-hoc tests highlighted that the difference in ratings between low and high fidelity of characters was present in the positive clip ($p < .0001$), but not in the negative clip.

- *Sadness Intensity:* A main effect of Fidelity was found ($F_{1,19} = 6.93$, $p = .016$) on sadness intensity ratings, where ratings were lower for low than for high fidelity characters ($p = 0.02$). A main effect was also found for Clip ($F_{1,19} = 214.38$, $p < .001$), where sadness ratings were higher in the negative compared to the positive clip.

4.1.3 Discussion: Experiment 2A

Our results indicate that state-of-the-art photorealism should be considered a positive choice for virtual human applications as it was found to be more attractive, less eerie, more human-like, and happier than lower fidelity depictions. Our results are in line with [Zibrek et al., 2019] where photorealism was found to be a positive choice in virtual reality, however other studies had conflicting results on the topic [Mäkäräinen et al., 2015, ?]. Additionally, previous work has shown that char-

acters with skin textures that are less detailed and show less blemishes and other skin irregularities are rated more attractive [Zell et al., 2015b], which is consistent with research in the cosmetics industry which shows that blurring skin texture can increase attractiveness[Fink and Matts, 2008]. Our low fidelity characters had much smoother skin textures with less detail than our high fidelity ones, however they were not rated more attractive, which we believe was due to the additional reduction in polygon detail counteracted any increase in attractiveness that the skin texture might have afforded.

In our study, our *low level of fidelity* condition was created by lowering the facial rig complexity, and the geometric and texture complexity to mimic characters currently seen across many applications, while retaining the same identity, motion, voice, lighting and rendering conditions as the high fidelity version.

While eeriness ratings were low for our high fidelity characters, it is possible that ratings might reach even lower levels with higher quality facial animation. We used real-time facial tracking and solving with only minimal clean-up in this study, and previous work has shown that lower quality facial animation can appear more eerie on realistic characters [McDonnell, 2012]. Future work should investigate the effect of facial tracking and solving.

Interestingly, we found that higher fidelity characters were rated to be happier when expressing a positive emotion. Previous work has shown that brighter lighting conditions can increase happiness ratings on virtual characters [Wisessing et al., 2020], however, our study is the first evidence that higher geometric and texture detail can also cause this effect, since our lighting was kept constant.

Our results also indicate that there was no effect of character appearance on uncanny ratings, which implies that future studies could perhaps use less variety in character appearance when studying these effects. We also found no effect of emotion on eeriness ratings, implying that future research might consider testing more neutral expressions to determine it would heighten uncanny ratings [Tinwell et al., 2011].

Finally, we were also interested in exploring whether variations in voice realism could be observed to have similar effects in related virtual character scenarios.

4.2 Experiment 2B: Synthetic Voices and Visual Consistency

Given that the aforementioned Metahuman characters represent the foremost quality of photorealistic virtual humans in circulation, our second experiment was designed to

combine different levels of voice realism, with higher photorealism. We expected that the greater salience in perceptual mismatch cues would have an influence on affinity and emotional responses. Specifically, we were interested in assessing whether an ambiguous voice might have the greatest influence on subjective responses. The voice realism conditions were established at three levels: Real, Synthetic and Text-To-Speech (TTS). As outlined below, the Real voice was a recorded human voice, and the TTS voice was generated entirely by a text-based speech generator. These TTS stimuli were written down in text and then transformed into non-human-like vocalisations by the speech generator, sharing no common prosodic features with the Real voice. The Synthetic voice combined both of these approaches; this stimulus was created by morphing the Real voice with a voice-tuning encoder system to reduce its pitch frequency and other prosodic features in order to make it sound less human. This was proposed to be a more ambiguous vocal stimulus compared to the Real and TTS voices.

We sought evidence that this stronger mismatch in voice and appearance realism could help to address two main hypotheses, namely

- *The synthetic voice conditions would affect the emotional response of participants to the character expressing different types of emotions.*
- *The synthetic voice would increase discomfort with the character.*

Previous methodologies in this domain offer support for the second hypothesis under heightened mismatch conditions, as outlined by Mitchell et al. [Mitchell et al., 2011b].

We also expected low scores for the social presence and realism measures used in the above design, given that the experiment was conducted on screen instead of in a virtual environment. This decision was forced by our goal of increased appearance realism, as the highest level of animation detail for the Metahuman characters (LOD 0) has been observed to induce reports of greater affinity than the highest available level for VR applications (LOD4).

As well as these considerations, a further feature of this study was an extra dimension of voice realism. Using a commercial Text-to-Speech system allowed us to expand the range of realism conditions for voice in the opposite direction to our appearance realism manipulation. In doing so, we also aimed to replicate and develop findings in similar experiments that have investigated the role of synthetic voices in virtual characters [Cabral et al., 2017].

Finally, This study was the beginning of a shift towards the study of empathy towards virtual characters in virtually mediated environments. There is a role to be

played by empathic responses in immersion and perspective taking of virtual characters [Ho and Ng, 2020]. Perspective-taking, also termed 'mentalising', is the act of attributing mental states to others, and considering the relationship of these states to the causal properties of the environmental scenario or context [van Loon et al., 2018a]. This perspective-taking response to emotional expressions has been termed 'Cognitive Empathy', and is one of the aspects of empathy measured in this study. The other two empathy dimensions that were explored are those of 'Affective Empathy' (the adoption of a character's expressed emotions by a perceiver) and 'Associative Empathy', which refers to the process of identifying or associating one's own experience with a character in a fictional setting [Shen, 2010].

Furthermore, an aspect commonly measured in research which studies the occurrence of the uncanny valley is that of 'familiarity', which has been observed to influence empathy towards digital humans expressing pain [Bouchard et al., 2013]. With the heightened photorealism for our second experiment, we aimed to explore whether the characters could induce empathy in participants.

4.2.1 Stimuli

The virtual human stimuli for our second experiment were developed in Unreal Engine 4.26[7] combined with the Metahuman Creator tool[8], and tracked at the highest available level of detail (LOD 0) from video of the actor performing three separate scenarios. The pipeline for generating these stimuli was the same for Experiment 1; as such, the higher animation quality was defined by a greater number of blendshapes (669) and joints (713) in the facial rig, as well as being highly geometrically complex (24000 facial vertices). Our aim was to investigate the effects on emotion and affinity at high levels of photorealism. The scenarios were recorded by our actor via a front-facing phone camera. We then used Faceware tracking software to retarget the performance to generate our Metahuman stimuli. (Figure 4.2). The three scenarios chosen for this experiment were intended to match those of [Zibrek et al., 2021]; Friendly, unfriendly and sad. In terms of emotional valence, this resulted in one positive and two negative valence scenarios.

There were three levels of voice realism used in this experiment. Firstly, we used the original natural voice of the actor whose motion was captured for generating the character animations. This was the designated 'Real' voice condition. The second level was created using the TANDEM–STRAIGHT [Kawahara et al., 2008, Kawahara et al., 2010] vocoder system outlined above, which was included for congruence with the previous work

[7]https://www.unrealengine.com
[8]https://metahuman.unrealengine.com

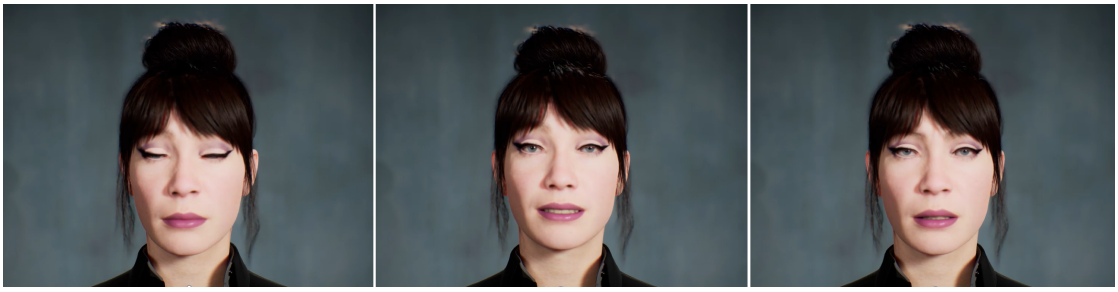


Figure 4.2: Screenshots from scenarios (left) Sad, (middle) Friendly, and (right) Unfriendly for character in Experiment 2. Participants viewed the character's performance with either a natural, synthetic or text-to-speech voice.

(This was the designated 'Synthetic' voice condition). In this experiment we transformed the VTL parameter, by the same factor of 1/5. The reason for not applying the transformation of the other speech parameters, as in the first experiment, was that for this stimuli it introduced excessive audible speech distortion, which could have an undesirable effect in the consistency of the results in regard of speech quality. From our own listening tests we found that the synthetic speech sounded sufficiently robotic/artificial compared to the 'Real' voice, which was our goal. By avoiding the transformation of the other parameters, particularly F0 and speech rate, we also aimed to preserve some expressive aspects of the Real voice. We also created voice clips from commercial Text-To-Speech software IBM Watson[9] which is the current industry standard for virtual agents and chatbots. The voice clips were created with Python at prosody calculated from the duration of the natural voice videos, and then manually mapped to the face movements. The motivation for inclusion of TTS vocals was twofold as it is more widely used in the industry than our previous synthetic voice condition and it also provided a lower level of vocal realism, which allowed us to expand the mismatch of perceptual cues when combined with greater appearance fidelity. This was the designated 'TTS' voice condition.

4.2.2 Measures

In keeping with the work outlined above, we included the same questionnaire validated by [Zibrek et al., 2019] for use in studies which focus on interactions with virtual characters, through a combination of established metrics for measuring social presence [Bailenson et al., 2003], emotional responses [Davis, 1983, Golan and Hill, 2006, Watson et al., 1988] and perceived realism [McDonnell et al., 2012] minus the place illusion item included in the work outlined above, given that we considered screen-based virtual character interactions insufficient for inducing the illusion of place. We also included a measure of empathic concern derived from related studies [Shen, 2010,

[9]https://www.ibm.com/watson

Ho and Ng, 2020] to assess whether the photorealism component of the stimuli could induce states of affective empathy, in addition to cognitive and associative state empathy, in participants outside of an immersive virtual environment (See Table 4.1, Empathy). Since the experimental design incorporated three levels of voice realism, we also used measures from [?] designed to acquire data on subjective impressions of synthetic voices in virtual character interactions (See Table 4.1, Voice Traits).

Table 4.1: Questions arranged by variable name. Each question could be answered on a scale from 1 – "Strongly Disagree" to 7– "Strongly Agree". Question with star were only asked in the *Avatar* condition.

Variable Name	Question
Concern	*"The character I just observed made me feel concerned.'*
Excitement	*"The character I just observed made me feel excited."*
Fear	*"The character I just observed made me feel afraid."*
Calm	*"The character I just observed made me feel calm, at ease."*
Appeal	*"I found the character appealing, likable."*
Eerie	*"I found the character eerie, creepy."*
Familiar	*"I found the character familiar, I have seen a similar person before."*
Overall Realism	*"I found the character realistic overall."*
Appearance Realism	*"I found the character's appearance realistic."*
Movement Realism	*"I found the character's movements realistic."*
Behaviour Realism	*"I found the character's behaviour realistic."*
Social Presence 1	*"It feels as if I am in the presence of another person in the room with me."*
Social Presence 2	*"It feels as if the character is watching me and is aware of my presence"*
Social Presence 3	*"The thought that the character isn't real crossed my mind often."*
Social Presence 4	*"The character appears to be alive."*
Social Presence 5	*"The character is only a computerized image, not a real person."*
Empathy 1	*"The character's emotions are genuine"*
Empathy 2	*"I can feel the character's emotions"*
Empathy 3	*"I can see the character's point of view."*
Empathy 4	*"The character's reaction to the situation is understandable."*
Empathy 5	*"When watching the video, I was fully absorbed."*
Empathy 6	*"I can identify the character in this message."*
Voice Likeability	*"How much did you like the character's voice."*
Voice Consistency	*"How well did you think the voice matched the character's appearance?"*
Voice Expressiveness	*"How expressive did you find the character's voice?"*
Voice Understanding	*"How easy did you find it to understand what the character was saying?"*

4.2.3 Participants and Procedure

Participants were recruited using Prolific[10], where we were able to attain an even split of participants on gender. The participants were provided with an information sheet, and consent form, a demographic survey and instructions before the experiment began. We provided a scenario of context for the virtual character interactions in order to mimic

[10]https://www.prolific.co/

the virtual environment in the study outlined above. The mixed factorial design that was employed meant that each participant was only exposed to one voice condition for three different scenarios. Each participant viewed the three scenario videos twice, for two blocks of questions. The first block consisted of our emotional response and empathy measures. The second block contained our items for social presence, realism and affinity ratings. The order of the videos was randomised between blocks, with three attention check questions per participant.

4.2.4 Analysis

Analysis was conducted on data acquired from the 60 participants recruited on Prolific, with a further 8 participants rejected on the grounds of failed attention checks. The 60 included responses were analyzed with a mixed factorial ANOVA on between subjects factor of 'Voice' and within subjects factor of 'Scenario', using IBM's statistical analysis software SPSS. Since several of our between-subjects measures violated homogeneity of variance on a Levene's test, we conducted a non-parametric Games-Howell test on the Bonferonni corrected data adjusted for multiple comparisons, to establish whether the effects of between subjects factor Voice were influenced by within subjects factor Scenario.

4.2.5 Results

We aimed to assess the influence of three levels of voice realism (Real, Synthetic, TTS) as our between subjects factor, with Scenario (Friendly, Unfriendly, Sad) as our within subjects factor. For a full account of the significant effects for this study, see Table 4.2. All unreported scales or effects were not significant.

H1: Voice type and social presence

As before, the 5 measured items of the Social Presence scale were tested for reliability for each Scenario, and due to sufficient correlation (Cronbach's alpha $\alpha = 0.82$ for Friendly, $\alpha = 0.80$ for Sad, $\alpha = 0.78$ for Unfriendly), we used a cumulative score of all 5 items (where items 3 and 5 were reversed) as the final result and treated it as a continuous scale.

A main effect for Voice was observed on the combined Social Presence variables for this experiment (See Table 4.2), in a deviation from the results of the first experiment. Particularly, the Text-To-Speech voice condition was measured as the lowest scoring on total means of our Social Presence metrics ($\overline{x} = 12.65$, $SD = 1.344$), implicating a lower level of voice realism evident in the TTS condition as a factor in subjective perceptions

Table 4.2: Summary of all main effects with corresponding post–hoc analysis for Experiment 2.

Effect	Analysis of Variance	Post-hoc
Realism		
MOVEMENT: Scenario	$F(2, 54) = 6.29, p = 0.003$	Sad was rated to have the most realistic movement compared to Unfriendly ($p = 0.008$)
BEHAVIOUR: Scenario	$F(2, 54) = 6.86$, $p = 0.002$	Behavioural realism was measured to be significantly higher in Sad compared to Unfriendly ($p = 0.005$).
BEHAVIOUR: Voice	$F(2, 54) = 3.382$, $p = 0.041$	Natural Voice more realistic than TTS Voice ($p = 0.04$).
Emotional Response		
CONCERN: Scenario	$F(2, 54) = 26.224, p < 0.001$	Concern for both Sad and Unfriendly measured to be greater than Friendly ($p < 0.001$).
EXCITED: Scenario	$F(2, 54) = 32.25, p < 0.001$	Friendly had higher ratings of excitement than both Sad and Unfriendly ($p < 0.001$).
EXCITED: Voice	$F(2, 54) = 8.19, p < 0.001$	Both Natural Voice ($p = 0.002$) and Synthetic Voice ($p = 0.003$) considered more exciting compared to TTS Voice .
AFRAID: Scenario	$F(2, 54) = 9.785, p < 0.001$	Afraid variable was reported greater in both Unfriendly and Sad compared to Friendly ($p < 0.001$).
CALM: Scenario	$F(2, 54) = 13.266, p < 0.001$	Friendly received more ratings on Calm than both Sad and Unfriendly ($p < 0.001$).
Voice Traits		
LIKEABILITY: Voice	$F(2, 54) = 12.642, p < 0.001$	Real voice was observed to be significantly more Likeable than either Synthetic or TTS voices($p < 0.001$) .
LIKEABILITY: Scenario	$F(2, 54) = 5.749, p = 0.004$	Unfriendly scenario was considered the least likeable compared to Friendly ($p < 0.007$) .
CONSISTENCY: Voice	$F(2, 54) = 4.969, p = 0.010$	Consistency ratings were lowest for Synthetic Voice compared to the Real voice ($p = 0.002$).
EXPRESSIVENESS: Voice	$F(2, 54) = 5.464, p = 0.007$	TTS was recorded as the least expressive compared to Real condition ($p = 0.010$).
Social Presence		
SOCIAL PRESENCE: Voice	$F(2, 54) = 4.58, p = 0.014$	TTS voice prompted the lowest social presence scores, significantly less than Synthetic Voice ($p = 0.017$).
Empathy		
ITEM 1: Scenario	$F(2, 54) = 5.38, p = 0.006$	Character's emotions perceived as most genuine for Sad scenario over Unfriendly ($p = 0.010$).
ITEM 2: Scenario	$F(2, 54) = 11.034, p < 0.001$	Sad scenario most induced participants to feel the character's emotions, in contrast to the Unfriendly condition ($p < 0.001$) and the Friendly condition ($p = 0.08$).
ITEM 5: Scenario	$F(2, 54) = 7.663, p < 0.001$	Subjects reported that they were most absorbed for Sad condition compared to bottom ranked Unfriendly scenario ($p = 0.014$) as well as Friendly scenario ($p = 0.011$).
Affinity		
APPEAL: Scenario	$F(2, 54) = 14.965, p < 0.001$	Friendly scenario rated significantly more appealing than Unfriendly ($p < 0.001$).
APPEAL: Voice	$F(2, 54) = 3.309, p = 0.044$	TTS condition was given significantly lower rating for appeal than Real Voice ($p = 0.043$).
FAMILIAR: Scenario	$F(2, 54) = 4.544, p = 0.013$	Friendly scenario was more familiar when compared to Unfriendly scenario ($p = 0.029$).
EERIE: Scenario	$F(2, 54) = 8.679, p =< 0.001$	Scores on Eeriness were strongest for the Unfriendly scenario compared to both Friendly and Sad ($p < 0.001$)

of social presence.

Conversely, the lowest observed means for Concerned ($\overline{x} = 3.15, SD = 0.247$) and Afraid ($\overline{x} = 2.65, SD = 0.230$) variables were recorded in the Friendly scenario, while Calm ($\overline{x} = 2.167, SD = 0.164$) was lowest for Unfriendly and Excited ($\overline{x} = 1.53, SD = 0.107$) was reported as the lowermost for the Sad scenario.

H3: Voice type and discomfort

Voice type was not observed to influence variables Eeriness or Familiarity, although a main effect was detected on Appeal ratings, with TTS voice condition evoking the least appeal ($\overline{x} = 2.983, SD = 0.270$) and Real Voice being the most appealing ($\overline{x} = 3.95, SD = 0.270$).

The virtual character was considered most appealing for the Real Voice condition in the Friendly Scenario ($\overline{x} = 4.550, SD = 0.400$) and least appealing for the TTS Voice in the Unfriendly Scenario ($\overline{x} = 2.1, SD = 0.329$).

Other results

Appeal variable was also significantly impacted by Scenario, with the character deemed to be least appealing in the Unfriendly Scenario overall ($\overline{x} = 2.750, SD = 0.190$). Similarly, an effect of Scenario was detected on both Eeriness and Familiarity. Eeriness variable was scored lowest overall for the Unfriendly Scenario ($\overline{x} = 4.883, SD = 0.205$) which mirrored the results for Familiarity ($\overline{x} = 3.1, SD = 0.227$).

Surprisingly, there were no effects observed of Voice type on our Empathy variable. However, Scenario did significantly effect empathy levels, particularly Item 1, Item 2 and Item 5. The Empathy variable was reported strongest on these measurements for the Sad condition, at its highest rating for this latter empathy measure ($\overline{x} = 5.317, SD = 0.231$).

For our Realism variables (See Table4.1) we observed a main effect of Scenario on Movement Realism at its strongest for the Sad Scenario ($\overline{x} = 3.467, SD = 0.219$) and Behaviour Realism also particularly high for Sad ($\overline{x} = 4.467, SD = 0.241$), with Voice only influencing Behaviour Realism (Figure **??**). The highest recorded Behaviour Realism mean was from the Natural Voice condition ($\overline{x} = 4.667, SD = 0.324$). There was no observed effect of Voice on Appearance or Overall Realism in the data analysis. Average means for Overall Realism ($\overline{x} = 3.733, SD = 0.400$) and Appearance Realism ($\overline{x} = 4.155, SD = 0.399$) across all Voice and Scenario Conditions indicate a general absence of observable effect.

Finally, for our subjective Voice Trait variables, we observed a main effect of Voice type on the first three measures Likeability, Consistency and Expressiveness. Interestingly,

the Synthetic voice was reported to be less likeable ($\overline{x} = 3.117, SD = 0.332$) and less consistent with appearance ($\overline{x} = 3.683, SD = 0.358$) than either the TTS or Real Voice. However, TTS Voice was considered to be less expressive overall ($\overline{x} = 3.550, SD = 0.353$). There was also a significant effect observed for Scenario on Likeability, with the character in the Unfriendly Scenario considered to have the least likeable voice ($\overline{x} = 3.417, SD = 0.225$).

H2: Voice type, traits and emotions

A significant main effect of Voice type was detected on variable Excited for all scenarios with the lowest scores recorded for the TTS voice condition for all scenarios ($\overline{x} = 1.5, SD = 0.171$) and the highest overall for the Real Voice ($\overline{x} = 2.4, SD = 0.171$). Observed interaction effects indicated that excitement for TTS condition was lowest for the Unfriendly Scenario ($\overline{x} = 1.20, SD = 0.250$) and the highest for the Synthetic Voice in the Friendly Scenario ($\overline{x} = 3.25, SD = 0.313$). There were no other effects of Voice observed in the data.

Scenario was observed to significantly effect all four emotional response variables; Concern, Excited, Afraid, and Calm, in stark contrast to Scenario results for the first experiment, but also predictably given the expressed valence of the respective scenarios. For instance, the Concern variable scored highest in the Sad scenario ($\overline{x} = 5.13, SD = 0.230$), while Afraid variable was most prominently rated in the Unfriendly scenario ($\overline{x} = 4.0, SD = 0.251$). Similarly, the Friendly scenario evoked stronger results on both Excited ($\overline{x} = 2.93, SD = 1.81$) and Calm ($\overline{x} = 3.68, SD = 0.241$).

4.2.6 Discussion: Experiment 2B

In terms of our voice mismatching methods, the TTS Voice was considered to be the least expressive voice condition, as well as the least appealing. While future work might be required to explore whether voice expressiveness impacts Social Presence with virtual characters, our finding that the photorealistic character with the TTS voice was the least appealing can be considered some evidence for our hypothesis on the effect of voice on uncanny ratings (H3), as well as a contribution to the growing number of perceptual mismatch theories which seek to explain the proposed Uncanny Valley effect. Additionally, the finding that the Real Voice condition was rated the most appealing could indicate that the photorealism of the virtual character was sufficiently perceptually aligned with a natural human voice. Our results mirror previous findings on preference for natural voices over synthetic voices, as described by [Cabral et al., 2017]. However, there was a noticeable novelty to our voice consistency results, as the synthetic voice was considered to be much less consistent with character appearance than the natural

voice.

The Voice Trait variables introduced in the latter study could shed some light on the first person experience of the Uncanny Valley effect. While the TTS condition was the least appealing on measures for uncanniness, the Synthetic voice was considered the least likeable and also least consistent with the appearance of the character, despite being more expressive than the TTS voice while there were no significant effects on Voice Understanding. The confusion may perhaps be shown in future work to arise at a level of habitability [Moore, 2016], wherein vocal agents are proposed to be most successful at speech when aligned with a users perceptually conforming interactional affordances. In other words, as applied to our study, with Synthetic Voice resembling the Real Voice much more closely than the TTS Voice, it was nonetheless belonging to a virtual human, albeit a photorealisic one. It nevertheless seems unclear to active perceivers what kind of interaction is occurring for the near natural voice, compared to the human or clearly non-human voices, potentially explaining our results for Voice Likeability and Voice Consistency, given that the TTS voice was specifically targeted at increasing inconsistency between voice and appearance realism in this study.

Our finding that the Real Voice condition was significantly both more appealing and more consistent with the appearance of the character is supported by previous work which has used human versus computer generated humanoid experiments to study voice and appearance combinations, and uncovered an overarching preference for consistency [Gong and Nass, 2007]. However, this is not mirrored in the lower voice realism condition, for the reasons described above. Our work found that our Synthetic voice was rated least likeable and least consistent with character appearance in our second experiment.The preservation of speech rate in the synthetic voice for Experiment 2 may have had an impact on these results for our consistency measure, and would be a useful further direction for research. Indeed, such work could be of use to both studies examining uncanny virtual humans of habitable vocal agents, as well as the combination of their respective traits in interactive contexts. Future work in this area could elucidate the extent of an influence of speech rate.

4.3 Experiment 2C: Animated Facial Expressions and Empathic Responses

This penultimate study was designed to assess how facial expressions of emotion could provoke different levels of empathy depending on the visual realism in virtual faces. Our final study, described in the next chapter, was intended to confirm the notion that

bimodal realism (appearance and voice) was the most promising avenue for emotion perception research in virtual characters. However, our due diligence required us to rule out the possibility that visual realism alone could be responsible for influencing emotion perception and empathic responses to the virtual characters we were employing as stimuli. As such, we designed a third virtual human experiment to test empathy for animated characters expressing emotion non-verbally, at different levels of visual realism. The emotional expressions were achieved using only facial expressions, as powered by the Facial Action Coding System (FACS).

The experiment included one between (*Visual Realism*: low, high) and two within participant independent variables (*Emotion*: happy, sad, neutral and *Character*: six levels which differed in shape, texture, gender and skin colour), observing their effects on the self reported variables of *empathy* (Cognitive Empathy, Affective Empathy and Compassionate Empathy) and *uncanny valley* (Appeal, Familiarity and Eeriness). Each of Cognitive, Affective and Compassionate empathy were measured with three-item subscales, while data was also gathered as related to trait empathy, as measured by the Interpersonal Reactivity Index [Davis, 1980]. Further details of these variables are described below.

4.3.1 Stimuli and Procedure

The character models used as stimuli in this study were the very same MetaHumans, as employed in the previous two studies. The difference between these stimuli and those used previously was that these characters were created using the Unreal Engine online application MetaHuman Creator[11]. Each MetaHuman is available in 8 levels of detail (LODs), with LOD0 having the highest level of visual realism and LOD7 having the lowest. For this experiment, LOD 5 was considered an appropriate choice for low realism, providing a noticeably less humanlike alternative to LOD0. The three emotional expressions were presented using preset Cubic Motion animations in the MetaHuman creator. Please see Supplemental Material for a video with stimuli examples.

For each gender, we used three MetaHumans chosen on three levels of character material variation, representing three different ethnicities. This decision was motivated by previous studies investigating the uncanny valley [Macdorman et al., 2008] which have ascribed low ratings of eeriness to photorealistic virtual faces but have outlined limitations of employing a single character. The character material and gender variation were also included to account for any potential biases in the participant sample, although there is conflicting evidence for such effects [Durupinar and Kim, 2022]

[11] https://metahuman.unrealengine.com



Figure 4.3: Our experiment investigated empathy towards a range of emotional virtual characters. Shown are two of the six virtual characters used in the study, at high visual realism (left) and low visual realism (right). Screenshots from the positive emotion animation are shown on top and negative on the bottom.

The measurements used for this study were derived from the Measure of State Empathy [Powell and Roberts, 2017], with three self-report items each for *Cognitive, Affective* and *Compassionate* Empathy. The respective items were summed to generate scores for each participant on each of these three dimensions, used in the analysis presented. This data was intended to capture the different kinds of empathy experienced by participants for each of the video stimuli that were presented on screen. We also included measurements for the Uncanny Valley as employed in previous similar studies [Diel et al., 2021], using the scales *Appeal, Familiarity* and *Eeriness*.

Participants were randomly assigned to one of the two realism conditions, with equal amounts of male and female participants assigned to each group. Upon following the link to the study, and prior to stimuli exposure, participants filled in sheets on informed consent and demographic information, which gathered information on age, gender and

video game experience. Additionally, they also gave information on their subjective interpersonal reactivity using the Interpersonal Reactivity Index [Davis, 1980], which was broken down into four separate dimensions of *Perspective Taking, Fantasizing, Empathic Concern* and *Personal Distress*. This inclusion in the demographics stage was intended to give our data a grounding in individual trait empathy.

For the stimuli exposure stage, participants were presented with clips of animated virtual humans that displayed positive, neutral or negative valence emotions. After each of the clips, participants were asked to rate perceived empathy and uncanniness for the virtual character, on slider scales measuring responses from 0-6 with twelve self-report items answered for each stimulus.

Since there were three types of character material and two character genders included with the differences in LOD and emotional valence, participants experienced 18 different clips, presented in randomized order. After experiencing all of the clips, participants were asked for feedback on their decision making processes via a text input box, to assess any technical difficulties that may have arisen during the experiment.

4.3.2 Results

The data were analysed using a 2x3x6 mixed analysis of variance (ANOVA), and was conducted on each dimension of the empathy and uncanny valley metrics to assess the influence of *Visual Realism* (low, high- between participants), *Emotion* (happy, sad, neutral- within participants) and *Character* (six levels-within participants) on the dependent variables described above. A mean score was derived from results for each of the three items for variables of Cognitive, Affective and Compassionate empathy. Before conducting our main analysis, the effect of character was tested to assess any evidence of bias in results based on character gender or ethnicity, with two different character genders and three different character ethnicities included in the stimuli. None of our results suggested an effect of character gender or ethnicity on dependent variables. Testing of assumptions highlighted that experiment data for all dependent variables of state empathy and the uncanny valley violated the assumption of sphericity ($p < 0.05$). We thus report Greenhouse-Geisser corrected ANOVA results.Additionally, Bonferroni adjustments were applied to post hoc tests to correct for multiple comparisons. Yet, due to the amount of analysis conducted, we advise that p values close to 0.05 be interpreted with caution.

Empathy metrics: There were no statistically significant main effects of visual realism on empathy scores ($p > 0.05$). A main effect of Emotion was recorded for all three dimensions of empathy (Cognitive Empathy: $F(2, 35) = 17.68, p < 0.001$; Affective

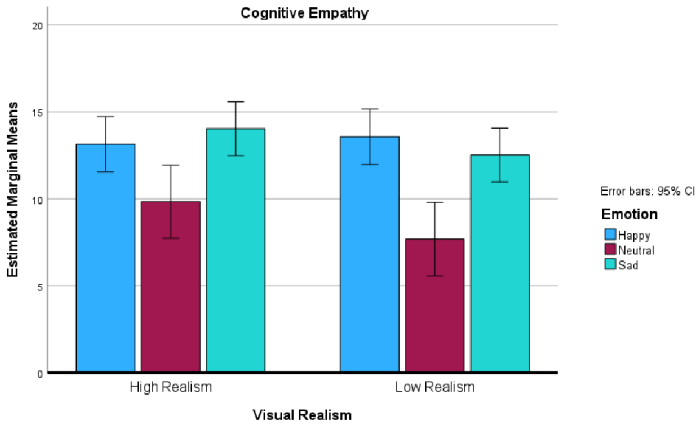


Figure 4.4: Estimated marginal means for Cognitive Empathy in all three emotion conditions

Empathy: $F(2, 35) = 6.82, p = 0.03$; Compassionate Empathy: $F(2, 35) = 21.41, p < 0.001$). Bonferroni corrected pairwise comparisons showed that for Cognitive Empathy the neutral condition ($\overline{x} = 8.76$, $SD = 5.46$, $p < 0.001$) led to significantly lower scores than the happy ($\overline{x} = 13.35$, $SD = 4.17$, $p < 0.001$) and sad conditions ($\overline{x} = 13.28$, $SD = 4.45$, $p < 0.001$). When compared to the neutral condition, ($\overline{x} = 5.40$, $SD = 4.90$, $p = 0.003$) Affective Empathy was significantly influenced by the happy emotion ($\overline{x} = 7.64$, $SD = 5.35$, $p = 0.007$). Similarly Compassionate Empathy was significantly influenced by the sad emotion ($\overline{x} = 10.46$, $SD = 5.07$, $p < 0.001$) when compared to neutral condition. All other main and interaction effects were not statistically significant ($p > 0.05$).

Uncanny Valley metrics: Surprisingly, Visual Realism across the conditions had no statistically significant effect on Appeal, Eeriness or Familiarity scores ($p > 0.05$). However, similar to the empathy metrics, a statistically significant main effect of Emotion was found for Appeal ($F(2, 35) = 7.88, p = 0.02$). Pairwise comparisons revealed that Appeal was particularly weak for the neutral condition ($\overline{x} = 3.14$, $SD = 1.80$, $p = 0.006$), compared to the happy ($\overline{x} = 3.71$, $SD = 1.80$, $p = 0.006$) or sad conditions ($\overline{x} = 3.30$, $SD = 1.77$, $p = 0.029$). We also found a statistically significant interaction effect between emotion and visual realism for the appeal subscale ($F(2, 35) = 3.85, p = 0.032$), with post hoc analysis showing that the lower level of realism led to higher reports of appeal for the happy emotion compared to neutral ($p < 0.001$) and sad conditions ($p = 0.037$). All other interaction effects were insignificant.

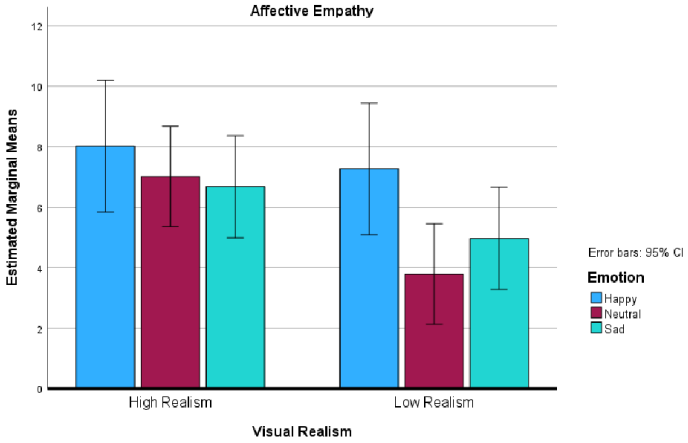


Figure 4.5: Estimated marginal means for Affective Empathy in all three emotion conditions

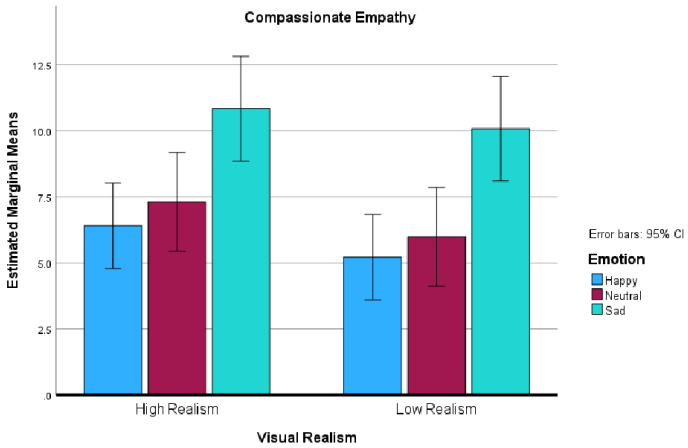


Figure 4.6: Estimated marginal means for Compassionate Empathy in all three emotion conditions

4.3.3 Discussion: Experiment 2C

This study was developed to examine whether virtual characters are capable of producing empathic responses, and whether these responses could be modulated by visual realism. Our main finding was that there were no significant differences on state empathy between visual realism conditions. Thus we suggest that visual realism alone, at least as operationalised in this experiment, does not lead to differences in empathic responses to virtual characters. This finding is useful for future works which explore empathy for virtual characters, as it suggests that researchers must assess other experimental paradigms and modes of realism. Although further studies are needed to replicate our findings, the absence of results in our experiment lends support to other works that fail to capture visual realism leading to strong differences in empathic response [Slater et al., 2006b, Lee et al., 2017].

Emotion had the greatest influence on empathic responses to virtual characters at both levels of realism, which is consistent with previous studies of virtual characters which have found scenarios to affect emotional responses more than realism variation [Zibrek et al., 2019]. Affective Empathy in particular was reported to be strongest in response to positive emotion expressions. Importantly, this result is consistent with work in human participants that has measured affective empathic responses to positively valenced treatments [Duan, 2000]. Similar results have been obtained from studies that have used virtual avatars to enhance positive affect [Oh et al., 2016]. Additionally, the reported lower scores in Cognitive Empathy during neutral stimulus exposure suggests that participants could not construct the impression of emotion from a face that was not clearly expressing a positive or negative valence.

It was surprising that that were no significant effects for Appeal, Eeriness and Familiarity, given that previous experiment 1A using these particular character models have revealed differences between perceived Uncanny Valley measurements for low and high fidelity animations. This contradiction may be due to the fact that the visual realism treatment was conducted between-subjects for this study, while previous work with these characters was conducted within-subjects. The results point to possible expectation effects as potentially influential in within subjects designs for measuring the uncanny valley [Lush et al., 2020].

Alternatively, previous research suggests that a mismatch of vocal and visual modes may play a role in perceived uncanniness [Mitchell et al., 2011a] which may also explain this deviation in results. Similar work has claimed that combinations of multi-modal realism conditions produce uncanny effects [Macdorman et al., 2008, Burleigh et al., 2013] more

reliably than the single variable of visual realism. To this end, our next experiment was designed to employ variations in both visual and vocal realism in these characters, in addition to positive and negative valence emotions.

Ultimately the results of the present study have shown that these visually realistic stimuli were not sufficient for inducing significant differences in uncanny or empathic responses between conditions. However, the results must be contextualised by certain limitations. As mentioned, the stimuli were presented with only one mode of realism, as there were no voice or behavioural realism conditions employed in this study. Similar constraints exist due to the the short duration of the stimuli clips, the non-lab based format, and the fact that the measurements used in this study relied on self-reported perceptions.

This research aimed to advance the literature further towards a focus on empathic responses to animated human representations, using a combination of standardised metrics and novel application. Future work could replicate this study in a virtual reality context, with an increasing role to be played by immersive environments on empathic responses [Slater et al., 2018, van Loon et al., 2018b]. The present study only focused on high and low levels of visual realism. More work involving different styles of characters, alternate modes of realism, and physiological or behavioral measures of empathy would be valuable in further developing virtual characters for emotional experiences.Further studies which seek to find alternative markers for empathic responses to realistic virtual humans will also add value to the work described in this paper.

4.4 Conclusions, intuitions and planned developments

The three experiments described in this section deal with questions of emotion and realism, with the focus on visual and vocal realism for all three experiments, separated by 1) visual realism variations with a real voice, 2) vocal realism variations in a realistic character and 3) visual realism variations in isolation.

In sum, there was evidence that emotion perception and character appeal could be influenced by visual realism in combination with a highly realistic voice. Similarly, lower levels of voice realism produced an appeal effect in combination with a highly realistic visual appearance. The third experiment narrowed the focus towards empathy, as a method for fine grained analysis of reactions to emotion perception, and showed that participants generally expressed empathy for virtual characters, regardless of variations in visual appearance when no voices were present in the stimuli.

These results provided us with motivation to further analyse the combination of visual and vocal affordances for emotion perception, and whether variations in their combination could influence both subjective reports and physiological responses to emotional virtual characters.

5 Experiment 3: Examining Empathic and Physiological Responses to Bimodal Realism in Emotionally Expressive Virtual Humans

The previous experiments, while each self-contained and providing relevant contributions to our understanding of virtual character perception, were ultimately useful for identifying a niche in the literature, that stood at the intersection of computer graphics, voice synthesis, cognitive science and human-computer interaction research. This niche may be defined by questions posed in the previous experiments; Since virtual characters at different levels of control, appearance and voice presentation are conducive to emotion communication, does virtual character realism afford emotion perception and hence, empathy? Similarly, since perceptions of virtual character embodiment, emotional intensity and realism can be measured to be influenced by individual subjective traits, are there subjective traits that can influence experiences of empathy at differing levels of realism?

The goal of this study was to assess this effect of visual and vocal realism on the physiological and self-reported responses of healthy participants during exposure to virtual humans expressing positive and negative valence emotions. The physiological measurements that were deployed here, namely eye-tracking and facial EMG, were proposed as a novel methodology for assessing empathy towards emotional virtual characters when combined with self reports of state empathy and measurements of trait empathy for each user. The measurements of empathy chosen for this experiment were the same as Experiment 2C; State Cognitive, Affective and Compassionate empathy were each measured with three-item subscales during the experiment, while trait Perspective-Taking, Empathic Concern, Personal Distress and Fantasising were measured for each participant

prior to the beginning of the experiment, using the Interpersonal Reactivity Index. While each of these physiological and self report instruments have been used in isolation for various scientific purposes, their combination in this way and in this domain of research is an additional contribution of this work.

As outlined in the background section, each of the physiological measurements have been used to focus on emotion perception and empathy. Below, hypotheses are described which are based on similar works that have developed comparable methodologies. Specifically, facial EMG focuses on measuring the electrical activity of two muscles (Zygomaticus major and Corrugator supercillii) which are proposed to exhibit increased activation under conditions of positive and negative emotional valence respectively. Additionally, eye-tracking methods are ubiquitous in human-computer interaction and cognitive science research, and have been useful for studying both real and virtual human interactions, as well as the perception of both faces and voices.

On the dimensions of emotion, the designated emotional valence conditions chosen for this experiment were specifically those of 'Happiness' and 'Sadness' respectively, given their mapping to positive and negative valence dimensions while maintaining neutral arousal. The terms 'Happy' and 'Sad' or 'Happiness' and 'Sadness' are used to refer to our emotional valence conditions from hereon, for the sake of brevity and clarity.

5.1 Hypotheses

The overall objective of the experiment is operationalized in detailed hypotheses, corresponding to the concepts described before. These hypotheses also reference the relevant analysis sections further in the paper.

Facial Electromyography and Emotion

Previous work has shown that there are increases in Zygomaticus major (abbreviated to Zyg) and Corrugator supercilii (abbreviated to Cor) activation in response to positive and negative valence emotions in real faces [Sato et al., 2008] and virtual faces [Ravaja et al., 2018]. We expected to replicate these results and additionally contribute with our novel hypothesis on realism, as measured by the EMG:

H1a: Cor activity will be greatest for highly realistic conditions of appearance and voice in the Sad condition. [Results 5.3.1]

H1b: Zyg activity will be greater for the highly realistic conditions of appearance and voice in the Happy condition. [Results 5.3.1]

H1c: High trait empathy will be correlated with increased Zyg activity in the Happy condition and Cor activity in the Sad condition. [Results 5.3.4]

Eye gaze and Empathy

Previous work with human face processing has suggested that emotional judgments of facial expressions are associated with increased eye gaze dwell time, while fixations on the mouth area are higher with speech processing tasks [Julie N. Buchan and Munhall, 2007, Vo et al., 2012]. Moreover, eye fixation behavior has also previously been associated with empathic traits [Cowan et al., 2014, Balconi and Canavesio, 2016]. In light of these findings, we hypothesized to find, thanks to the ET method that:

H2a: Overall, eye fixation will be greatest for the eye-region of the virtual faces for highly realistic conditions of voice and appearance. [Results reported in 5.3.2]

H2b: High trait empathy will correlate with greater eye-region fixations. [Results reported in 5.3.4]

H2c: High state empathy will correlate with greater eye-region fixations. [Results reported in 5.3.4]

State Empathy and Realism

In terms of empathy, we were looking for evidence that state empathy reports could be observed to show significant variance between levels of realism. Given that reports of emotion intensity can vary between realism conditions as measured in experiment 2A, we expected that empathy levels would also be subject to change depending on whether the characters were exhibiting higher, lower levels of realism. While open to whether the UV may influence this result, we hypothesized that:

H3: State empathy would be higher for the highly realistic appearance and voice condition [Results reported in 5.3.3]

Uncanny Valley

Our study also seeks to expand research into the UV effect, with a view to furthering our understanding of the underlying emotional processes. This focus follows from experiment 2A, where these characters yielded uncanny ratings that are greater for the lower realism characters. Similarly, the mismatch hypothesis [Mitchell et al., 2011b, MacDorman and Chattopadhyay, 2016] further informed our hypotheses on the UV. With this in mind, we also expected that, given the findings of [Mäkäräinen et al., 2015], although motor mimicry for sadness could potentially be measured in the data, there

remained a possibility that the UV could interfere with mimicry behavior in the zygomaticus major muscle:

H4a: The highly realistic conditions would evoke the lower UV ratings and higher ratings of appeal, realism and familiarity. [Results reported in 5.3.3]

H4b: The mixed realism and the low realism conditions would be considered more uncanny, with lower ratings of appeal, realism and familiarity. [Results reported in 5.3.3]

5.2 Materials and Methods

The present study was developed and conducted in accordance with the Declaration of Helsinki and the Trinity College Guidelines for Good Research Practice. The study protocol (reference no: 2913) was approved by the Research Ethics Application Management system (REAMs) of Trinity College Dublin (REC: School of Computer Science and Statistics) on 14/11/2023. Before commencing the experiment, participants read the information sheet and signed the informed consent form. Participants were compensated with a 15 euro gift card at the end of the experiment. Below, we outline the virtual human stimuli creation pipeline, measurement methods and procedure.

5.2.1 Stimuli Synthesis

Voice

The realism of the voice (high or low) was one of the variables explored in this study. The high realism voices were real male and female actors performing emotional expressions of happiness and sadness, as used in experiment 2A. The clips were originally recorded via phone and cleaned in Audacity 3.3.3 to normalize volume and reduce background noise. For the voice with low realism, the same actor scripts were used to generate synthetic speech using a high quality speech synthesiser from VoiceTune.ai. The synthesised voices preserved qualities of humanness and emotion that were intended to be present for each condition. However, the goal was to produce voices that were not as realistic as recordings of human actors. As such, the chosen voices were synthesised at 22 kHz and then upsampled to 48 kHz, to match the original audio recordings of the actors. While maintaining some qualities of human voices that intonate emotion, the synthetic voices were significantly less realistic than the actor recordings, as defined by the smoother pitch frequencies and absence of breaths. This technology was developed from the basis of the Tandem-straight pipeline that was used in the previous voice experiment that was outlined in experiment 2B.



Figure 5.1: Two of the character models used in the study, expressing happiness (top) and sadness (bottom) at both high (left) and low (right) levels of visual realism.

Characters' appearance

The visual aspects of the stimuli were created in Unreal Engine 4.26 (UE4.26)[1]. These stimuli took the form of Metahuman character models created using the Unreal Engine online application MetaHuman Creator[2]. Each MetaHuman is available in 8 levels of detail (LoD), with LoD0 having the highest level of visual realism and LoD7 having the lowest. The differences in fidelity arise from differences in geometric complexity (for example, 24000 facial vertices at LoD0 versus 1300 at LoD4), texture complexity, and the use of strand-based hair (LoD0) versus card-based hair (LoD4). Each MetaHuman has a full facial and body rig for animation, with the complexity of these rigs also varying across the different realism. As mentioned in the previous chapter, the LoD0 facial rig consists of 713 joints and 669 blendshapes, whereas the LoD4 facial rig has only 84 joints and no blendshapes. An identical rigging setup across MetaHumans means that a given animation can be faithfully reproduced on multiple MetaHumans. In our case, real-time rendered animations were captured of actors performing our positive and negative valence conditions. The valence expressions were derived from literature outlining particular phrases that may be used to represent positive or negative valence emotions, as described in experiment 2A. The major change for these stimuli compared to those used in experiment 2A was a change in animation style. For the present experiment, each recorded performance was then tracked using an offline solver, applied to the animations by CubicMotion (a leading provider of automated performance-driven animation to the games industry) in order to create naturalistic animated expressions using industry standard performance tracking software[3].This provided us with 2x male (happy/sad) and 2x female (happy/sad) animations. These animations were then mapped to the MetaHuman facial animation rig and the resulting animation was recorded using UE4.26's Sequence Recorder feature. The commonality of the facial rig across MetaHumans meant that for each emotional valence a single recorded animation could be used across all female characters and similarly for the male characters. The different components of the stimuli (MetaHuman, animation, lighting, virtual camera, and audio) were pieced together using UE4.26's Sequencer feature. The same virtual camera setup and standard three-point lighting setup were used for all stimuli. For each gender and emotional valence, the same audio track and animation were used, meaning the only difference between stimuli was the actual MetaHuman itself. Several post-processing effects such as ray-tracing and anti-aliasing were used on the final stimuli which were rendered using the UE4.26 Movie Render Pipeline. Each stimulus was rendered at 30 FPS and at a resolution of

[1] https://www.unrealengine.com

[2] https://metahuman.unrealengine.com

[3] $https://northedge.com/our_portfolio/cubic-motion/$

2048 $\times$ 1536 and downsampled to 1024 $\times$ 768. An Nvidia GeForce RTX™ 3090 GPU was used for the rendering. For each gender, we used two MetaHumans chosen on three levels of character ethnicity, with two males and two females (one white male, one asian male, one black female, one asian female). This decision was motivated by previous studies investigating the UV [Macdorman et al., 2008] which have ascribed low ratings of eeriness to photorealistic virtual faces but have described the limitation of using only a single style of possible human variation. The character ethnicity and gender variation were also included to account for any potential biases in the participant sample. We used these six MetaHumans with LoD0 for our high fidelity condition and LoD4 for our low fidelity condition. For our purposes, LoD4 was considered an appropriate choice for low fidelity, representing human likenesses at a considerably lower level of quality compared to LoD0, in terms of both polygon count and appearance fidelity. LoD4 (Figure 5.1 (right)) is also considerably less humanlike than LoD0 (left).

5.2.2 Participants

Recruitment of participants was conducted via college mailing lists and on-campus posters. Initially, 40 subjects were recruited but 1 was excluded because of poor vision. Therefore, 39 healthy young adults (20 males, 19 females; age: mean = 26.81 years, SD= 5.19 years; education: mean= 18.52 years, SD= 2.49 years; ethnicity: 14 Asian, 1 Black, 19 White, 4 Others; video-game experience: 8 Expert, 15 Moderately competent, 12 Sporadic users, 3 Never used) were recruited from the local community of students in Dublin (Ireland).

They all were self-reported as native English speakers or fluent in English, to have no history of neurological, psychiatric, or motor disorders, and normal or corrected to normal vision. The sample size was estimated using G*Power 3.1 (Universitat Kiel, Germany) with an a priori power analysis for an F test with repeated measures, within factors. Considering the EMG results as the main outcome, previous studies with similar procedures found small to moderate effect size [Golland et al., 2018]. Setting the same effect size (f(V) = .3), we calculated a total sample size of 39 participants (with the α error probability set at .05 and the power set at .95).

5.2.3 Measures

Electromyography (EMG)

Facial electromyographic (EMG) activity of two muscles, Zygomaticus Major and Corrugator Supercilii, was recorded on the left side [Rukavina S, 2016], according to the guidelines for facial EMG placement [Fridlund AJ, 1986]. We used a Trigno Research



Figure 5.2: Electrode placement for measurement of Zygomaticus major (bottom) and Corrugator Supercilii (top) activity

System (Delsys, USA) connected via Bluetooth (wireless) to a dual-head EMG (Trigno Duo Sensor): each head (25 x 12 x 7 mm, 21 g) is equipped with two bipolar AG surface electrodes (10 mm apart), attached on the skin with a provided sticker. A specific software (Trigno Discover, Delsys, USA) recorded and locally stored the EMG data, which were then pre-processed with EMGworks Analysis 1.0 (Delsys, USA). The automatic sampling rate was 1259 Hz throughout the experiment. The EMG data were recorded only during the videos (static and dynamic phases) for a total of 86 seconds for each block.

Eye-Tracking (ET)

Gaze data (fixation duration and fixation count) was gathered using a Tobii Pro Spectrum eye tracker (Sweden) and an EIZO FlexScan EV2451 monitor (Japan) with a pixel resolution of 1920×1080 (52.8×29.7cm). The monitor was positioned above the ET bar which stood approximately 9cm off the desk. Given that the recommended distance between the ET hardware and the participant's eyes ranged from 55-75cm, the settings for the calibration stage using Tobii Pro Eye-Tracker Manager 2.6.0 were set at 65 cm, to allow for 10 cm of adjustment to distance during each phase of calibration. Gaze data were recorded at a sampling rate of 1200 Hz only during the videos (for a total of 86



Figure 5.3: Pilot run of participant setup for eye-tracking measurement. The Tobii bar was placed below the stimuli which were presented on screen.

seconds for each block). The Areas of Interest (AOIs) designed for the face were drawn using the Tobii tracker AOI tool (Figure 5.7). Previous work has shown that greater size and variation in AOI production generates significant noise, so our method was chosen to highlight the relevant facial features as specifically as possible [Hessels et al., 2016], while adopting styles that have been used to generate AOIs in previous face-gaze studies [Auyeung et al., 2015, Vehlen et al., 2021].

Questionnaire

The questionnaire used in this study was necessary to capture subjective aspects of the experience. The same questionnaire was repeated at the end of each condition (block); the order of items was randomized at every repetition. Participants were asked to report their level of agreement with each statement on a 1 ("totally disagree") to 7 ("totally agree") Likert scale. The items on the questionnaire were used previously in experiment 2A and 2C.The state empathy measures from [Powell and Roberts, 2017] are intended to capture subjective reports of cognitive empathy, affective empathy and compassionate empathy.

Uncanny Valley metrics have been the subject of repeated development [Ho and MacDorman, 2010a, Ho and MacDorman, 2017]. In attempting to measure the effect, familiarity has widely been used as a measure of realism, although recent work has suggested that a direct realism scale is important to include [Diel et al., 2022]. See Table 5.1 for the actual items, their measure, and descriptive statistics.

Interpersonal Reactivity Index (IRI)

The Interpersonal Reactivity Index [Davis et al., 1980] is a commonly used survey instrument that contains four separate subscales measuring different dimensions of trait empathy (Empathic Concern, Perspective Taking, Personal Distress and Fantasising). For instance, the Empathic Concern subscale may be considered a stable measurement of trait emotional empathy, while the Perspective Taking subscale can be understood as trait cognitive empathy [Cowan et al., 2014]. While there are contentions that Personal Distress cannot be considered a standalone measure of empathy [Pulos et al., 2004], it was included in our analysis in order to capture any possible effects. This measurement was to broaden our empathy focus. While the questionnaire at the end of each block was intended to capture empathy for each character's condition, it was important to contextualize any such reports by recording empathic traits that individuals displayed to greater or lesser degrees. Therefore, the IRI was assessed once for participant at the very beginning of the experiment. The subscales were examined to assess internal reliability for each of Perspective Taking ($\alpha = .698$) Empathic Concern ($\alpha = .790$) Personal Distress ($\alpha = .615$) and Fantasising ($\alpha = .803$) For each subscale, descriptive statistics are reported for the entire sample (average, SD) and for each subgroup (N, average, SD). The four subscales defined by the IRI were assessed independently: Perspective Taking (M = 16.95, SD = 3.33), Empathic Concern (M = 19.92, SD = 4.75), Personal Distress (M = 12.87, SD = 3.92), and Fantasizing (M = 18.87, SD = 4.88). Of particular interest were the Perspective Taking and Empathic Concern scales, given their associations with Cognitive and Affective empathy [Jolien Van der Graaff and Branje, 2016].

5.2.4 Procedure

Participants were invited into the experiment room, where they were seated in front of the screen and accompanying ET device. They were asked to fill out a demographics questionnaire assessing video game experience, education, age, gender, and ethnicity. They next completed the Interpersonal Reactivity Index questionnaire for measuring trait empathy [Davis et al., 1980] (see 5.2.3 for details).

Following this introduction, the experimenter explained the procedure for the EMG and ET data collection, and the EMG electrodes were placed on the subject's face: participants were told that the EMG would be measuring electrical activity naturally produced in the Zyg and Cor muscles in the face, although they were not informed about the reasons for the measurements. The emotional reactivity component of the EMG and ET data were unknown to participants until after the experiment, in order to minimize expectation confounds. After the electrodes were attached to the face, the participant

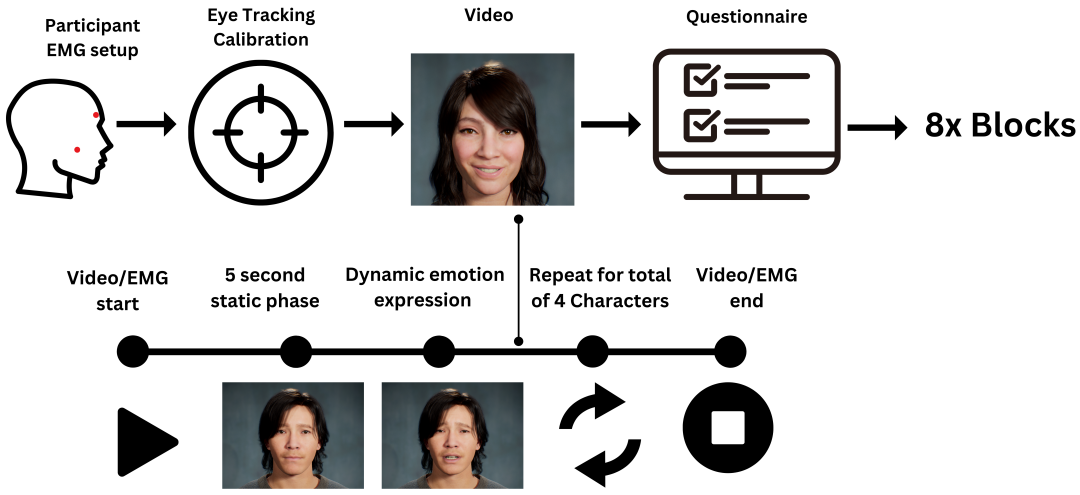


Figure 5.4: Schematic of experiment procedure.

was directed to voluntarily activate the muscles to calibrate to the EMG data collection software and to check the correct positioning of the electrodes over the muscles.

The next stage was explained to participants as a recurring three phase cycle corresponding to Eye-tracking calibration - videos - questionnaire, for a total of eight blocks (see Figure 5.4 for schematic of this procedure), corresponding to the eight conditions combining the variables realism (High or Low) for appearance and vocal, and emotion (Happy and Sad). After putting on the studio headphones, they were told that these eight blocks would begin (the entire experiment lasted about 45 minutes). First, participants were calibrated to the ET device placed below the screen, which involved a two stage process of calibrating head position, gaze direction and drift (lasting about 3 minutes, see 5.2.3 for details). After the calibration phase, participants were exposed to four videos in a randomized order of virtual characters. Videos were split into static and dynamic phases: each video contained the first 5 seconds of inanimate faces (hereinafter, static phase), followed by 16.5 seconds of emotional expression (hereinafter, dynamic phase), for a total of 21.5 seconds. Therefore, the block with four videos ran for a total of 86 seconds. The 5 second static phase was chosen to facilitate analysis for both the EMG data and the Eye tracking data. It was used as a mechanism for resetting eye behaviour between dynamic phases, and hence was removed from the ET analysis. For the EMG, it provided us with a resting phase window to noise correct the rest of the recorded signal for the dynamic phase. The EMG recording started at the same time as the first video in each block (see 5.2.3 for details).

Trial Blocks

The video stimuli were characterized by virtual humans expressing positive (happy) and negative (sad) emotions, at two levels of appearance realism (see 5.2.1 for details), and two levels of voice realism (see 5.2.1 for details), which ultimately composed the experimental conditions, corresponding to the eight blocks (the first letter corresponds to the appearance realism, the second letter to the vocal realism and then either the happy or sad emotion):

- A_hV_hHAP: high appearance realism, high vocal realism, Happy.
- A_hV_hSAD: high appearance realism, high vocal realism, Sad.
- A_lV_hHAP: low appearance realism, high vocal realism, Happy.
- A_lV_hSAD: low appearance realism, high vocal realism, Sad.
- A_hV_lHAP: high appearance realism, low vocal realism, Happy.
- A_hV_lSAD: high appearance realism, low vocal realism, Sad.
- A_lV_lHAP: low appearance realism, low vocal realism, Happy.
- A_lV_lSAD: low appearance realism, low vocal realism, Sad.

At the end of each block of videos, participants completed a questionnaire measuring state empathy and uncanny valley responses. The same block, composed of calibration - videos - questionnaire was then repeated. The order of conditions, and therefore blocks, were randomized between participants, with all participants experiencing all conditions for a within-subjects design.

5.3 Results

Analysis was conducted with MATLAB R2023b (The MathWorks Inc., USA) and with SPSS Statistics V28 (IBM Corp, USA), Statistica 7.0 (TIBCO Software Inc., USA), and with R 4.3.2 (ARTool, emmeans and PMCMRplus packages, The R Foundation, USA). Normality was assessed using a Shapiro–Wilk test. All values displayed are (average (standard deviation)). When significant comparisons were detected, we reported the corresponding analysis results, in addition to effect sizes (appropriate to the conducted statistics). Each hypothesis is referenced in the relevant results outlined below.

Measure	Statement	A_bV_bHAP	A_bV_bSAD	A_bV_iHAP	A_bV_iSAD	A_iV_bHAP	A_iV_bSAD	A_iV_iHAP	A_iV_iSAD
Cognitive Empathy	I understood how the characters were feeling	5.359 (1.112)	5.462 (0.913)	4.744 (1.390)	4.615 (1.616)	5.051 (1.169)	5.513 (1.167)	4.667 (1.691)	4.590 (1.482)
Cognitive Empathy	I knew what the characters felt emotionally	5.205 (1.301)	5.462 (0.942)	4.846 (1.387)	4.513 (1.520)	5.000 (1.298)	5.385 (1.161)	4.641 (1.513)	4.615 (1.515)
Cognitive Empathy	I could identify the feelings the characters were having	5.487 (1.048)	5.769 (0.872)	5.077 (1.222)	4.897 (1.518)	5.128 (1.239)	5.564 (1.373)	4.949 (1.395)	4.718 (1.685)
Affective Empathy	The characters' feelings transferred to me	4.051 (1.572)	3.923 (1.306)	3.179 (1.587)	2.872 (1.559)	3.282 (1.638)	3.410 (1.650)	3.026 (1.597)	2.872 (1.525)
Affective Empathy	I felt the same way as the characters	3.615 (1.600)	3.692 (1.490)	3.256 (1.482)	3.051 (1.572)	3.154 (1.631)	3.513 (1.715)	2.923 (1.579)	2.897 (1.603)
Affective Empathy	I experienced the same emotions as the characters	3.846 (1.514)	3.949 (1.486)	3.051 (1.521)	3.103 (1.619)	3.179 (1.485)	3.487 (1.652)	2.846 (1.548)	2.974 (1.614)
Compassionate Empathy	I had feelings of concern for the characters	2.821 (1.745)	4.385 (1.462)	2.487 (1.467)	3.564 (1.744)	2.538 (1.466)	3.718 (1.638)	2.333 (1.325)	3.103 (1.667)
Compassionate Empathy	I experienced feelings of sympathy towards the characters	3.949 (1.538)	4.897 (1.119)	3.179 (1.554)	3.769 (1.677)	3.231 (1.646)	4.410 (1.464)	2.974 (1.646)	3.462 (1.502)
Compassionate Empathy	I felt a sense of compassion for the characters	4.026 (1.547)	4.667 (1.325)	3.308 (1.398)	3.718 (1.669)	3.590 (1.551)	4.308 (1.507)	3.128 (1.436)	3.333 (1.562)
UV	I found the characters appealing, likeable	4.692 (1.472)	4.667 (1.383)	3.974 (1.495)	3.564 (1.465)	4.128 (1.576)	4.282 (1.521)	3.462 (1.430)	3.256 (1.499)
UV	I found the characters eerie, creepy	2.974 (1.739)	2.718 (1.486)	3.692 (1.809)	3.795 (1.750)	3.385 (1.711)	3.308 (1.608)	4.077 (1.692)	4.282 (1.761)
UV	I found the characters familiar, I have seen similar characters before	4.615 (1.695)	4.846 (1.631)	4.615 (1.771)	4.179 (1.715)	4.821 (1.604)	4.692 (1.673)	4.128 (1.894)	4.205 (1.976)
UV	I found the characters realistic overall	5.000 (1.338)	5.051 (1.146)	3.949 (1.555)	3.385 (1.632)	3.974 (1.678)	4.359 (1.564)	3.154 (1.479)	2.897 (1.518)

Table 5.1: Questionnaire assessed at the end of each condition (block). The first columns represent the measures underlying each statement of the questionnaire (not disclosed to participants); the third column describes the actual statement that participants rated on a 1-7 Likert scale (not ART transformed used in the analysis). The last eight columns represent the averaged (SD) values between participants for each condition.

5.3.1 EMG Hypotheses (H1 a-c)

Data were preprocessed and resampled at 1000 Hz. We then designed Notch filters at 50 Hz and 100 Hz, setting the bottom end cutoff point of 20 Hz with the top end of the frequency set at half the sampling rate. Thus the frequency range of interest to this work was between 20-500 Hz. These Notch filters [Sato et al., 2008, Ravaja et al., 2018] were specifically applied at 50 Hz to reduce the influence of common power grid noise, and at 100hz to compensate for harmonics. The Notch filters were chosen to eliminate possible noise, with the relatively high cutoff frequency chosen to balance the noise reduction with the actual muscle activation, so as to maximize the chance of gleaning useful activation data and minimize the chance that noise would influence the results. Along with the filtration process, we calculated the Root Mean Square (RMS) for each participant at each of the eight conditions, to acquire a single value RMS for each participant, muscle, and condition. This calculation provided an average score for muscle activity in mV in both the Zyg and Cor across each period of stimulus exposure where independent variables were manipulated. As such, we acquired sixteen different RMS values per participant, eight for each muscle, which were then analyzed for variance across conditions of Appearance, Voice and Emotion. Notably, the 5 seconds of static phases at the beginning of each video per condition were removed from the analysis. Moreover, the RMS values were noise corrected against 3 seconds of the first static phase (2-5sec) at the beginning of the block, before the onset of the dynamic emotional stimulus. The preceding two seconds were not analyzed. RMS for each muscle was analyzed separately to assess variations in activation between conditions, realism, and emotions. The distribution of the data from each muscle was found to be not normally distributed, and thus we applied an Aligned Rank Transform (ART) to the data [Lisa A. Elkin, 2021], [Simon Kimmel and Boll., 2023]. We then conducted a 2x2x2 repeated measures ANOVA for each muscle separately comparing the eight conditions and the related factors: realism for Appearance (High, Low), realism for Voice (High, Low),

	A_hV_hHAP	A_hV_hSAD	A_lV_hHAP	A_lV_hSAD	A_hV_lHAP	A_hV_lSAD	A_lV_lHAP	A_lV_lSAD
Cor	.0035 (.0019)	.0039 (.0021)	.0034(.0020)	.0038(.0022)	.0039 (.0027)	.0039 (.0021)	.0036 (.0015)	.0039 (.0021)
Zyg	.0067 (.0198)	.0022 (.0009)	.0031 (.0039)	.0024 (.0012)	.00446 (.0115)	.0027 (.0017)	.0057 (.0183)	.0025 (.0015)

Table 5.2: Descriptive statistics of the RMS recorded with the EMG. The first line refers to the conditions, the second and third are the RMS values for corrugator supercilii (Cor) and zygomaticus major (Zyg) respectively. Values are reported as average (SD).

and Emotion (Happy, Sad). When a significant effect was found, post hoc analyses were conducted with Tukey corrected pairwise comparisons.

Results

A statistically significant main effect was found for the Zygomaticus Major muscle for the factor of Emotion ($F_{1,39} = 10.07, p = .002, \eta 2 = .33$), where Happy emotion provoked greater activation in the Zyg compared to the Sad Emotion.

There were also significant interactions between Appearance and Voice ($F_{1,39} = 4.37, p = .038, \eta 2 = .12$), and Appearance, Voice and Emotion ($F_{1,39} = 10.25, p = .002, \eta 2 = .27$). Tukey's pairwise comparisons between conditions revealed these interactions to be specific to condition A_hV_hHAP, where zygomaticus activity was significantly greater compared to conditions A_hV_hSAD ($p < .001$), A_hV_lSAD ($p = .034$), A_lV_hHAP ($p = .049$), A_lV_hSAD ($p < .001$) and A_lV_lSAD ($p = .002$).

There was a significant main effect recorded for factor of Emotion on Corrugator Supercilii activity ($F_{1,39} = 10.41, p = .001, \eta 2 = .63$), where the Sad Emotion had significantly greater Cor activity compared to Happy. There were no interactions found in the Cor data. See Table 5.2 for descriptive statistics. (See Hypotheses 5.1).

5.3.2 ET Hypotheses (H2 a-c)

The ET recorded fixations for the three AOIs (Eyes, Mouth and Nose). Eye fixations were measured on two dimensions: the amount of time spent fixating on an area (fixation duration) and the number of fixations on a given area (fixation count). In the process of exporting the data, we removed each of the 5-second static phases from the analysis, which were then handled separately from the dynamic phases. The data obtained from the dynamic phases of the four videos in each condition were then averaged. Dynamic phase data for fixation duration and count were each analyzed separately. They violated the Shapiro-Wilk test for normality, and hence the data were adjusted using an Aligned rank transformation (ART) to account for the non-normal distribution of data. Following the transformation model, a four-way repeated measures ANOVA (2x2x2x3) was conducted comparing between factors realism for Appearance (High, Low), realism

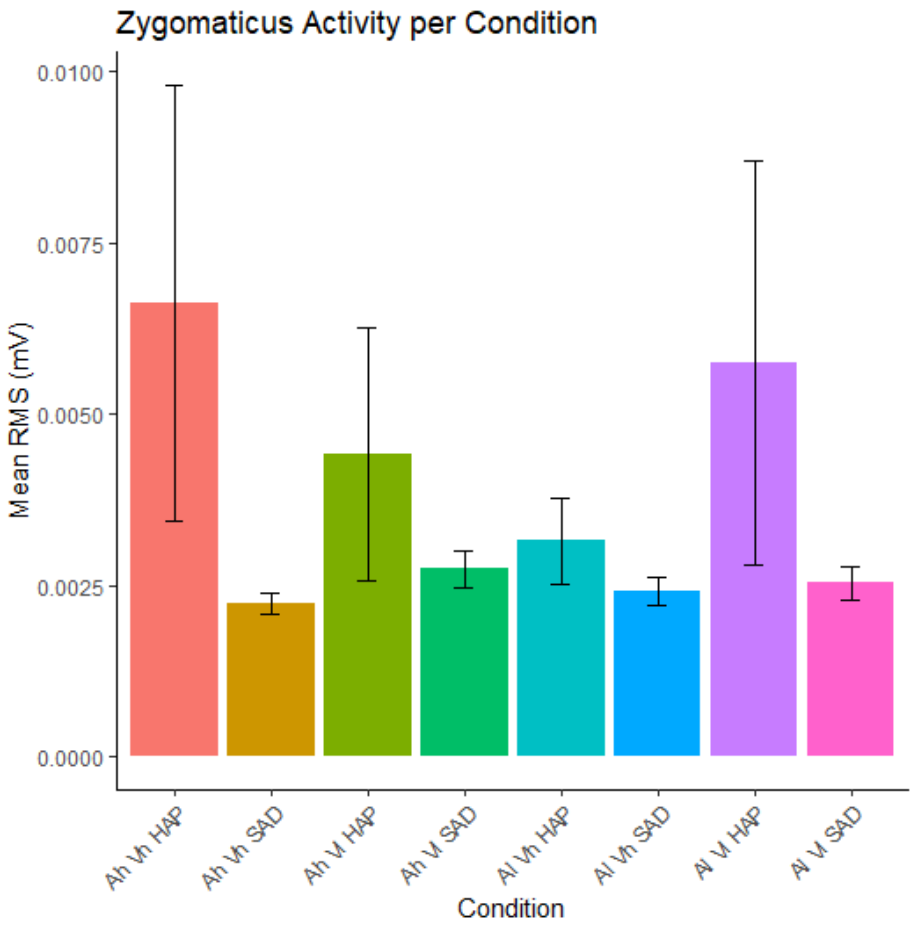


Figure 5.5: Bar plot of mean RMS for Zygomaticus activity across all conditions, with standard error bars.



Figure 5.6: Bar plot of mean RMS for Corrugator activity across Happy (left) and Sad (right) conditions, including standard error bars.

for Voice (High, Low), Emotion (Happy, Sad), and AOI (eyes, mouth, nose). Post hoc comparisons were performed with Tukey corrected pairwise comparisons are reported too.

Results

Eye fixations for dynamic phases were observed to vary statistically significantly in terms of both fixation duration and count.

For the fixation duration, a significant main effect was found for Areas of Interest (AOI) ($F_{1,39} = 703.91, p < .001, \eta 2 = .99$). Pairwise comparisons indicated that participants spent a significantly longer amount of time fixating on the eyes of the virtual characters, regardless of realism or emotion ($p < .001$) compared to mouth ($p < .001$) and nose ($p < .001$). There was no recorded influence of realism or emotion on fixation durations for any AOI.

Similar results were observed on fixation count ($F_{1,39} = 131.96, p < .001, \eta 2 = .96$). See Table 5.3 for descriptive statistics. Once again, the fixation count on the eye region of the virtual characters was significantly greater compared to mouth ($p < .001$) and nose ($p < .001$).

A main effect of Emotion was also recorded for fixation count ($F_{1,39} = 10.44, p =$



Figure 5.7: ET areas of interest (AOIs) shown on high appearance realism character (white male), eyes (highlighted in pink), nose (highlighted in yellow), and mouth (highlighted in orange)

.

.001, $\eta 2 = 0.01$), with a higher number of fixations for the happy emotion on all AOIs than the sad emotion. See Table 5.4 for descriptive statistics. (See Hypothesis 5.1 and also Figure 5.8).

5.3.3 State Empathy and Uncanny Valley Hypotheses (H3, H4a-b)

In order to examine subjective experience following the stimuli, we analyzed each empathy category (i.e. Cognitive, Affective and Compassionate Empathy) and Uncanny Valley item of the questionnaire separately. The distribution of each item was not normally distributed, and thus we applied an Aligned Rank Transform to the data [Lisa A. Elkin, 2021, Simon Kimmel and Boll., 2023]. We then conducted a 2x2x2 repeated measures ANOVA to assess the effect of realism of Appearance (High, Low), realism of Voice (High, Low), and Emotion (Happy, Sad) on the transformed questionnaire item scores. Post hoc comparisons were performed with Tukey's contrasts. In our

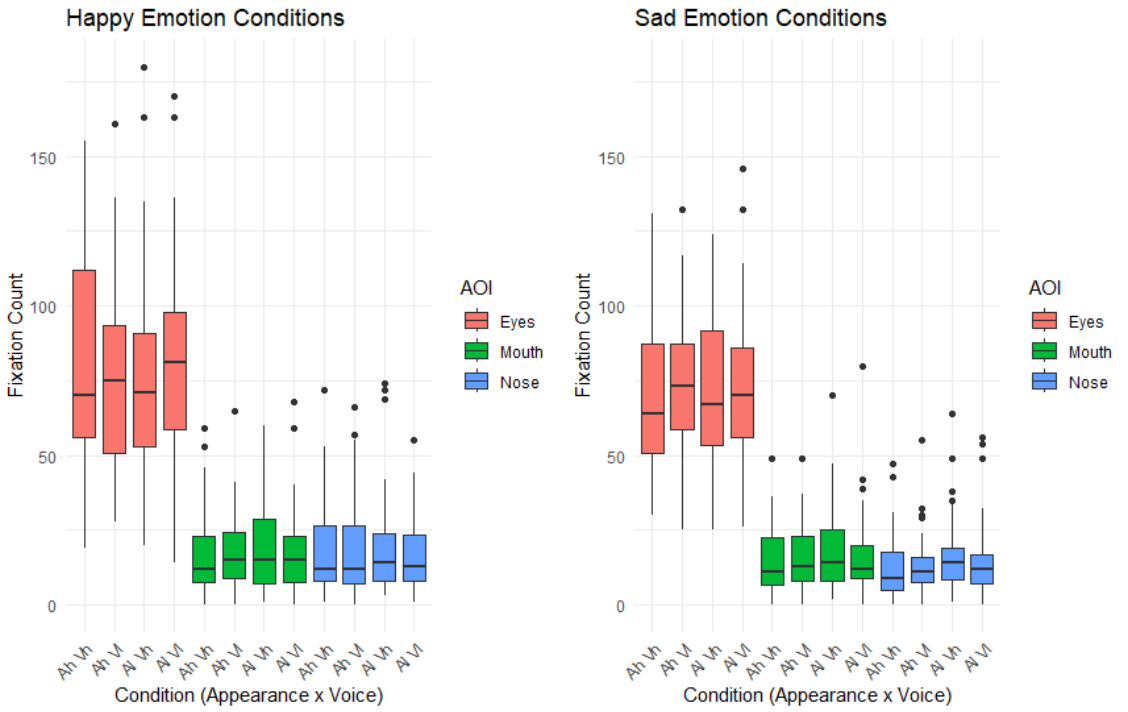


Figure 5.8: Box plots for eye fixation count across Happy and Sad emotion conditions, for all Areas of Interest. Significantly higher fixation count was observed on all Areas of Interest for Happy (left) compared to Sad (Right).

data, state empathy metrics demonstrated strong internal reliability for Cognitive Empathy ($\alpha = .924$) Affective Empathy ($\alpha = .913$) and Compassionate Empathy ($\alpha = .841$)

See Table 5.1 for the statements of the questionnaire and the descriptive statistics.

State Empathy Measures - Cognitive, Affective and Compassionate Empathy (H3)

In terms of state empathy, a main effect was found for Voice on all three dependent variables. Main effects were recorded for the effects of Voice on Cognitive Empathy ($F_{1,39} = 26.33, p < .001, \eta 2 = .70$), Affective Empathy ($F_{1,39} = 29.20, p < .001, \eta 2 = .62$), and Compassionate Empathy ($F_{1,39} = 45.42, p < .001, \eta 2 = .32$), indicating that the real voice evoked higher reports of Cognitive, Affective and Compassionate empathic responses compared to the synthetic voice.

Appearance realism was recorded to influence two of the three dimensions of empathy, namely Affective ($F_{1,39} = 12.26, p < .001, \eta 2 = .27$) and Compassionate ($F_{1,39} = 14.94, p < .001, \eta 2 = .11$) empathy, showing that the high realism appearance was rated higher than low realism for both Affective, and Compassionate empathy.

A main effect of Emotion was observed for Compassionate Empathy ($F_{1,39} = 63.82, p <$

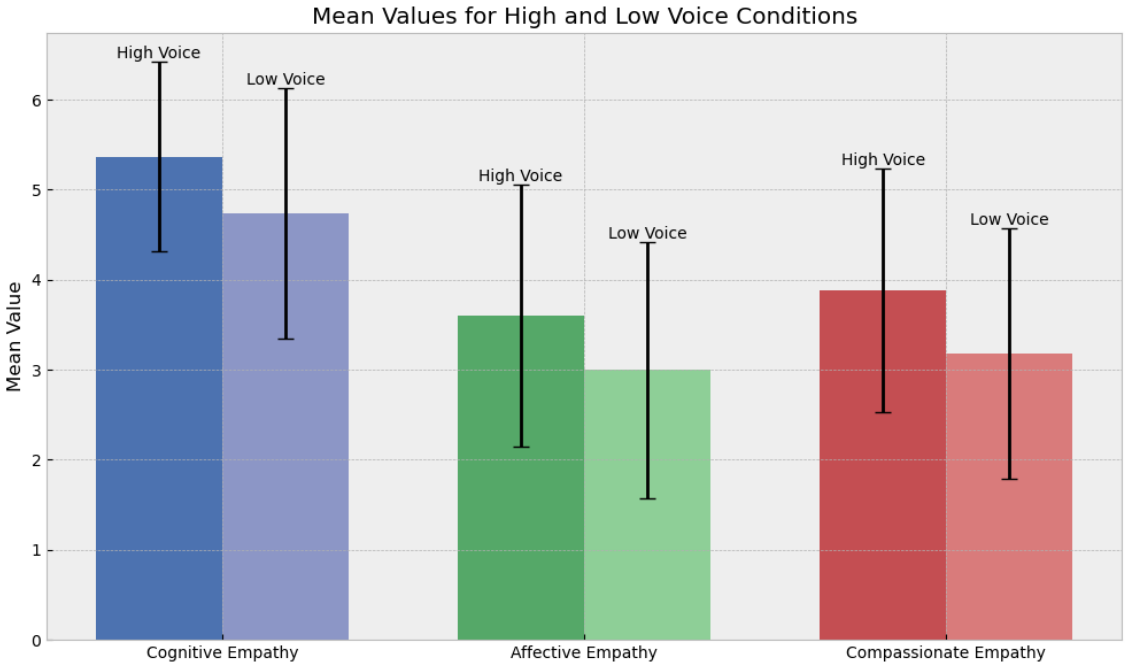


Figure 5.9: Bar plots for reported state empathy across voice conditions.

.001, $\eta 2 = .44$) where the Sad condition evoked higher reports of Compassion compared to Happy.

Uncanny Valley Measures (H4 a and b)

Main effects were observed on our appeal measure for Appearance ($F_{1,39} = 16.75, p < .001, \eta 2 = .21$), with the realistic appearance rated significantly more appealing. There was also an effect of Voice ($F_{1,39} = 58.22, p < .001, \eta 2 = .71$), with the real voice also considered more appealing. Also, there was a main effect of Voice on reports of eeriness ($F_{1,39} = 33.89, p < .001, \eta 2 = .49$) and realism ($F_{1,39} = 124.48, p < .001, \eta 2 = .66$) with the synthetic voice considered to be significantly more eerie and less realistic. There was also a main effect for Appearance on both eeriness ($F_{1,39} = 8.94, p = .003, \eta 2 = .13$) and realism ($F_{1,39} = 34.65, p < .001, \eta 2 = .21$), with the less realistic appearance considered more eerie and less realistic. See (5.1).

5.3.4 Exploratory Analysis

Additional findings are described below. Spearman correlations (r and p are reported, N is always 39) were run to assess co-arising phenomena within the dataset. We were interested to see whether there would be any observable relationship between muscle activity or eye fixation with trait empathy (measured in the IRI) and questionnaire data on state empathy and Uncanny Valley items. These analyses were conducted with a

view to either supporting or clarifying previous hypothesised results. We only describe the significant correlations between measurements recorded within the same condition (for example, questionnaire and EMG within A_hV_hHAP).

Correlations between State Empathy, Uncanny Valley, and EMG

Two significant correlations were recorded for Corrugator activity; in condition A_hV_hHAP, there was a positive correlation for mean RMS and Familiarity ($r = .321, p = .039$), while in the condition A_lV_lSAD, there was a positive correlation recorded for Cor RMS and Compassionate Empathy ($r = .321, p = .046$).

Correlations between State Empathy, Uncanny Valley, and ET

In the A_hV_hHAP condition, fixation count on the Mouth area was positively correlated with reports of character Appeal ($r = .323, p = .045$). For the condition A_hV_hSAD, there was a positive correlation between Eye area fixations and Affective Empathy ($r = .365, p = .022$) as well as a negative correlation between reports of Eeriness and Eye area fixations ($r = -.327, p = .042$). Additionally, in the condition A_hV_lSAD, there were positive correlation for Affective Empathy with Mouth area and ($r = .442, p = .005$) and Nose area ($r = .448, p = .004$), as well as well as character Appeal ($r = .392, p = .019$) for Mouth area fixations.

In terms of fixation durations, there were significant results for condition A_hV_lSAD, where there were positive correlations between Compassionate Empathy and Mouth area dwell time ($r = .364, p = .023$) as well as appeal ($r = .335, p = .037$), and a negative correlation between Eye area fixation durations and reports of Affective Empathy ($r = -.396, p = .012$).

Correlations involving trait empathy

We finally conducted correlations to assess an the relationship between dependent variables and empathic traits, measured by the IRI.

There were significant correlations observed between trait Fantasising and Eye movement behaviour. There was a negative relationship recorded between the Fantasy scale and fixation count on the mouth area for the A_lV_hNeg condition ($r = -.402, p = .01$).

Fantasising was also positively correlated with increased zygomaticus major activity in the A_hV_hHAP condition ($r = .481, p = .005$) and the A_lV_lHAP condition ($r = .322, p = .05$). In the A_hV_hHAP condition, Fantasising was positively correlated with Cognitive Empathy ($r = .334, p = .037$) (See 5.1).

	A_hV_hHAP	A_hV_hSAD	A_lV_hHAP	A_lV_hSAD	A_hV_lHAP	A_hV_lSAD	A_lV_lHAP	A_lV_lSAD
eyes	35.737 (12.649)	34.169 (8.963)	34.174 (14.486)	34.341 (10.593)	33.837 (13.269)	36.357 (12.126)	34.854 (11.949)	35.117 (11.81)
mouth	9.186 (8.243)	8.808 (7.219)	10.813 (10.359)	10.003 (7.824)	10.505 (9.591)	10.799 (10.033)	10.295 (9.261)	9.624 (7.979)
nose	7.722 (7.924)	5.082 (4.449)	8.256 (8.183)	7.752 (6.760)	)7.003 (6.725)	5.348 (4.253)	7.088 (5.910)	6.544 (5.984)

Table 5.3: Descriptive statistics of the fixation durations recorded with the ET, expressed in seconds. The first line refers to the conditions, the second, third and fourth are the values for the three AOIs, eyes, mouth and nose, respectively. Values are reported as average (SD).

	A_hV_hHAP	A_hV_hSAD	A_lV_hHAP	A_lV_hSAD	A_hV_lHAP	A_hV_lSAD	A_lV_lHAP	A_lV_lSAD
eyes	80.359 (34.218)	70.179 (25.802)	76.154 (36.869)	72.949 (26.575)	76.846 (30.643)	73.179 (25.594)	81.590 (33.378)	75.359 (25.728)
mouth	17.436 (14.184)	14.487 (11.116)	19.333 (15.844)	17.590 (14.221)	18.795 (13.784)	16.333 (11.437)	17.513 (15.030)	16.615 (14.807)
nose	18.692 (16.392)	13.051 (11.133)	20.179 (17.790)	17.641 (13.897)	18.051 (16.203)	13.538 (10.802)	17.359 (13.435)	15.103 (13.591)

Table 5.4: Descriptive statistics of the fixation count recorded with the ET, expressed in n. The first line refers to the conditions, the second, third and fourth are the values for the three AOIs, eyes, mouth and nose, respectively. Values are reported as average (SD).

Perspective Taking was negatively correlated with Compassionate Empathy for the condition A_hV_lSAD ($r = -.377, p = .018$), as well as measures of character familiarity ($r = -.317, p = .049$) and realism ($r = -.375, p = .019$). For the condition A_lV_hSAD there was an observed negative correlation between Perspective Taking and perceived Eeriness ($r = -.49, p = .029$).

Finally, in the A_hV_hHAP condition, Empathic Concern was positively correlated with Cognitive Empathy ($r = .343, p = .033$).

5.3.5 Cross-Experimental Analysis

In order to fully understand the observed effects, a further analysis was conducted between experimental data and reports of cognitive, affective and compassionate empathy. The data added to this analysis came from our third virtual human experiment, where facial expressions of emotion were presented to participants without the use of vocal expressions. This analysis comes with an important caveat; the stimuli used for Experiment 3 in the previous section were different to the visual-vocal stimuli used in the present study. While we used the same metahuman characters at two levels of visual realism, the participants that are added to this dataset experienced no voice exposure, and thus the characters presented on the screen were animated differently to our visual-vocal emotional stimuli. Nonetheless, we consider this to be a more natural non-voice emotion expression compared to simply removing the voice from our speaking stimuli, which we anticipated may create a sense of eeriness that could interfere with the data.

The same ART model was used to analyze the data with an additional 40 participants empathy reports added to the dataset. The resulting dataset of 79 participants was

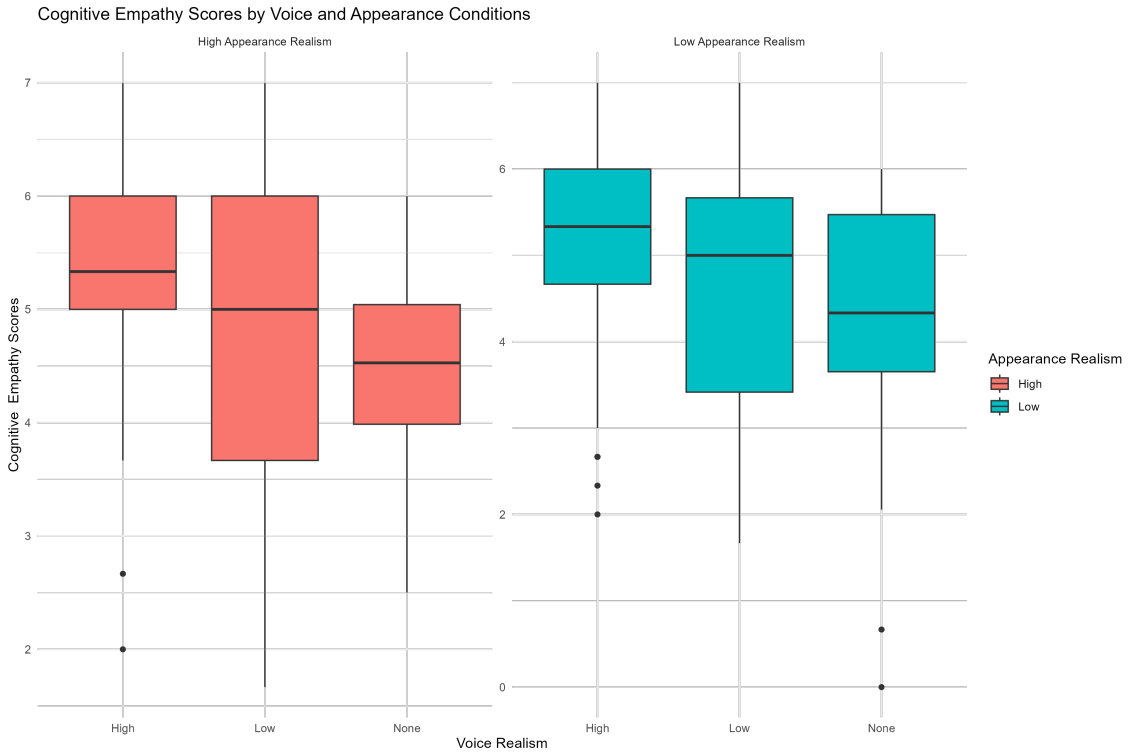


Figure 5.10: Box plots for reported cognitive empathy across voice and appearance conditions.

analysed using a 2x3x2 repeated measures ANOVA, to understand how adding vocal affordances for emotion to an expressive virtual character may impact perceived levels of empathy.

Upon processing and analysing the additional data, a pattern consistent with previous results was observed. Participants were significantly swayed by the presence of a Voice in terms of Cognitive Empathy ($F_{1,79} = 33.17, p < .001, \eta 2 = .92$) with pairwise comparisons suggesting that the real voice evoked significantly more cognitive empathy than the synthetic voice ($p < .001$) and the no voice condition ($p < .001$) with the synthetic voice also provoking higher reports of cognitive empathy compared to the no voice condition ($p < .001$).

In terms of Affective Empathy, the previous result was maintained in terms of Appearance, with high appearance realism provoking more Affective Empathy compared to low ($F_{1,79} = 12.19, p < .001, \eta 2 = .12$). Moreover, there was a significant effect of Voice ($F_{1,79} = 30.98, p < .001, \eta 2 = .74$), with pairwise comparisons indicating that the real voice was responsible for significantly more affective empathy than the synthetic voice ($p < .001$) and the no voice condition ($p < .001$) with the synthetic voice also considered to provide higher reports of affective empathy compared to the no voice condition

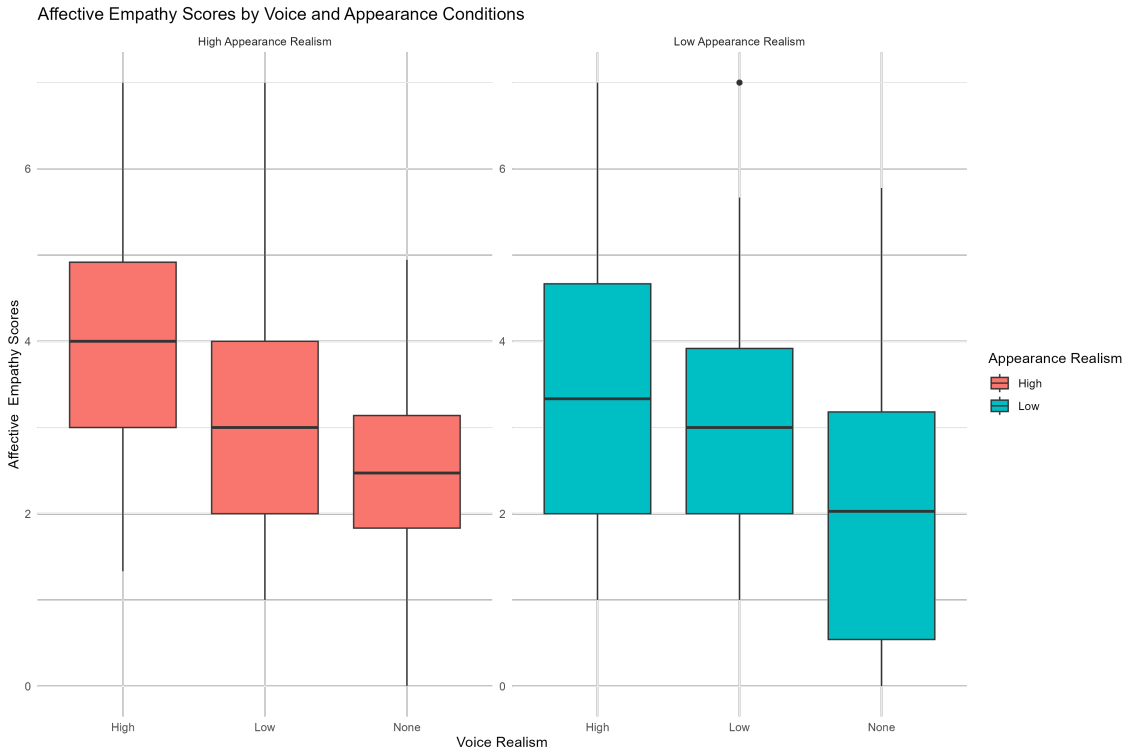


Figure 5.11: Box plots for reported affective empathy across voice and appearance conditions.

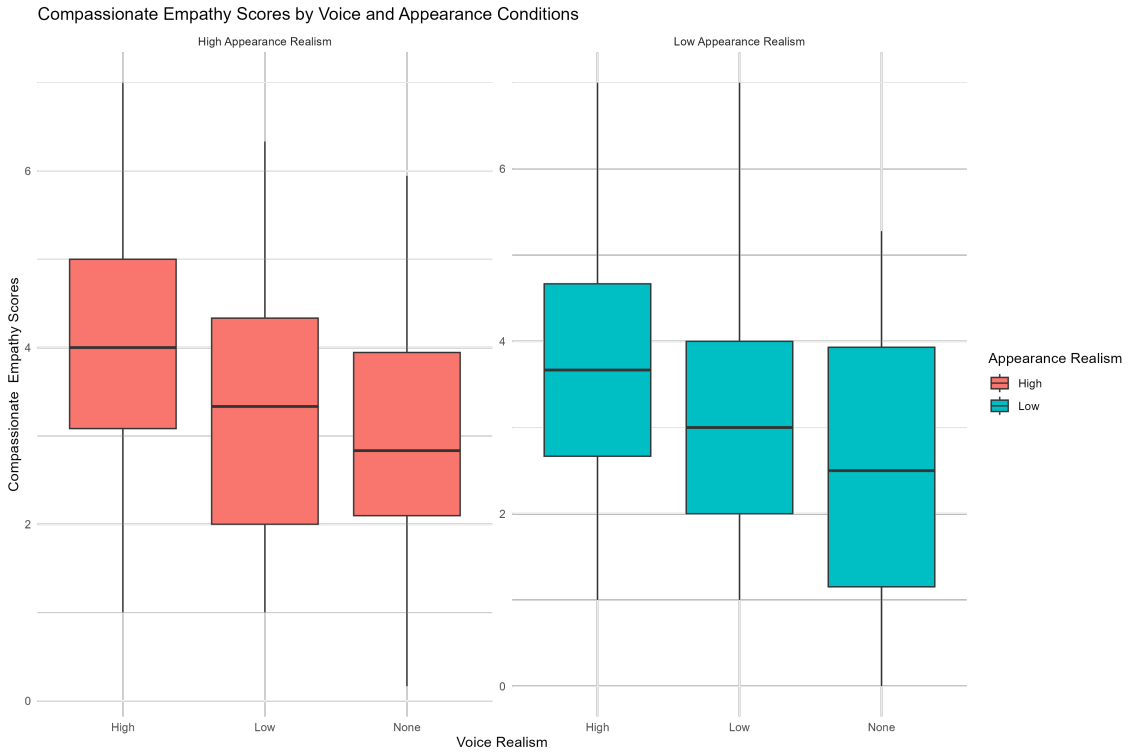


Figure 5.12: Box plots for reported compassionate empathy across voice and appearance conditions.

($p < .001$).

Finally, with the inclusion of the extra data, Compassionate Empathy, was significantly influenced by Voice ($F_{1,79} = 43.22, p < .001, \eta 2 = .13$), with participants reporting more compassionate empathy compared to the synthetic voice ($p < .001$) and the no voice condition ($p < .001$), with synthetic voice also evoking higher reports of compassionate empathy compared to the no voice condition ($p < .001$). Moreover, there was a main effect recorded for Appearance, with high appearance realism provoking for compassionate empathy compared to low ($F_{1,79} = 13.79, p < .001, \eta 2 = .02$). Finally, the effect of Emotion from the previous analysis was also maintained, with participants reporting more compassionate empathy for the sad conditions compared to the happy conditions ($F_{1,79} = 109.81, p < .001, \eta 2 = .83$).

5.4 Discussion

The main purpose of this study was to understand whether bimodal realism could influence experiences of empathy at a physiological and subjective level.

There is some evidence to suggest an effect of realism on facial EMG responses, but this finding is mediated by emotion. Specifically, zygomaticus activity for the most realistic happy condition was significantly higher compared to most of the other conditions.

Additionally, it is clear that the emotional valence had a strong effect on physiological responses. For the sad emotion, corrugator activity was measured to be significantly greater for the sad emotion compared to happy. This emotion effect is also highlighted by the significantly higher zygomaticus activity for happy emotion conditions. These findings suggest an occurrence of motor empathy, in the form of emotional mimicry.

Realism also had an impact on subjective reports of empathic responses, as high visual and vocal realism provoked significantly higher state empathy reports.

An additional important finding of this work was that trait Fantasising correlated with zygomaticus activity in positively valenced conditions. This correlation was specific to conditionswith the highly realistic voice and appearance combination, or low realistic voice and appearance combination, and not observed in conditions of mismatched voice and appearance realism.

Finally, in terms of the Uncanny Valley, unrealistic appearance and voice determined more eeriness and lower perceived realism.

With regards to the hypotheses outlined in the background section, the results provide

both surprising and expected insights.

H1a: *Cor activity will be greatest for highly realistic conditions of appearance and voice in the Sad condition.* 5.3.1

H1b: *Zyg activity will be greater for the highly realistic conditions of appearance and voice in the Happy condition.* 5.3.1

H1c: *High trait empathy will be correlated with increased Zyg activity in the Happy condition and Cor activity in the Sad condition.* 5.3.4

There is some support for both H1a and H1b. Our results suggest that, in relation to H1a, we cannot confirm the hypothesis, as there was no recorded effect of realism on Cor activity, but there was an effect of Sad emotion. However, EMG results appear to provide evidence in significant support of hypothesis H1b, as Zyg activity was highest for the conditions of Happy Emotion, and there was a three way interaction between Appearance Realism, Voice Realism and Emotion that was also measured to influence activity in this muscle.

With regards to the emotion effect, this phenomenon has been discussed in the context of emotional mimicry as a feature of emotion perception [Chartrand and Bargh, 1999, Künecke et al., 2014], whereby spontaneous facial mimicry in emotional states can occur within 1000 ms of exposure. This would seem to avoid the requirement for constructing an internal representation when perceiving an emotion [Moody et al., 2007]. Conversely, there is evidence that the act of inferring emotion from expressions enhances the occurrence of spontaneous facial mimicry [Murata et al., 2016], which would imply a role of cognitive representations in emotion perception. While these theoretical distinctions remain debated, this study adds to the literature on emotional mimicry by demonstrating its occurrence under both positively and negatively valenced conditions during interactions with virtual characters. Our findings contrast with those of other studies which have suggested that positive emotion is more readily mimicked than negative in human-human emotional scenarios, inferred from findings on the Zyg muscle [Rymarczyk et al., 2016]. The effect of emotion on Cor activity also supports previous results on responses to virtual characters expressing emotions [Ravaja et al., 2018].

While reports of state empathy (discussed below) suggest that there are affordances offered by the real human voice and realistic animated representation that allow for empathy, there also appears to be evidence that our reactions to realism in characters may be observed at the physiological level, given the interaction effect of Appearance Voice and Emotion on the Zyg activity data. Further confirmatory studies should be conducted to examine this point, as the most realistic positive valence condition of

A_hV_hHAP appeared to provoke significantly greater Zyg activity compared to five of the other conditions. This finding may provide a useful direction for studying how different facial or vocal qualities, in both humans and virtual characters, provide affordances for emotion perception. In this study, it appears that the affordances offered by the highly realistic happy character invite users to react with congruent facial muscle activity. This may be viewed as a form of social smile response [Schilbach et al., 2005], with the implication being that appearance and voice realism, modulated by positive valence, appear to enhance this facial reactivity to expressions of happiness.

Results also provide partial confirmation of H1c. There is evidence that facial mimicry of emotion has previously been linked to empathy [Holland et al., 2021], which it has been in this study, by the finding that participants who scored highly on empathic trait Fantasising exhibited greater Zyg activity in positive valence conditions, specifically those of A_hV_hHAP and A_lV_lHAP. This correlation implies a relationship between matched realism modes and greater Zyg activation for trait empathisers. Given previous work has found mismatched realism to be a causal factor in the uncanny valley [Mitchell et al., 2011b], there is an implication here that the empathic trait of fantasising may heighten sensitivity to mismatches in appearance and voice.

H2a: *Overall, eye fixation will be greatest for the eye-region of the virtual faces for highly realistic conditions of voice and appearance.* 5.3.2

H2b: *High trait empathy will correlate with greater eye-region fixations.* 5.3.2

H2c: *High state empathy will correlate with greater eye-region fixations.* 5.3.4

While our findings support other studies which find a similar bias towards the eyes for computer-generated characters [Raiturkar et al.,], as well as eye contact in making emotional judgements for real faces [Julie N. Buchan and Munhall, 2007], there was no effect of realism observed in the eye tracking data. As such there is only partial support for H1a. There was no direct association found between eye-gaze and trait empathy, with evidence providing neither support nor refutation of H2b since there was high eye-gaze dwell-time overall. However, contrary to previous works eith real human stimuli [Cowan et al., 2014, Balconi and Canavesio, 2016], there was no recorded difference in eye gaze behaviour that could be correlated with trait empathy. The general tendency towards eye gaze dwell time that was observed in our data reflects the proposition that eye contact in virtual character interactions is important for achieving joint attention in dyadic and collaborative settings [Schilbach et al., 2005].

An additional finding from the eye tracking data appears to augment the EMG findings on the occurrence of facial mimicry, as there was an overall effect of positive valence

on the number of eye fixations for all AOIs. This finding appears to suggest that the scanning for visible joint attention affordances was moderated by the perception of emotion in the character. In other words, participants were more likely to explore visual affordances under positive valence conditions. This could be due to the possibility of smiling behaviour proving important for join attention mechanisms, similarly to eye gaze dwell time, but may also be a contributor to the occurrence of emotion-specific facial mimicry [Schilbach et al., 2005].

There is some support for H2c, specific to the A_hV_hSAD condition and reports of state Affective Empathy correlating with eye area fixations. Interestingly, there is also evidence against H2c in the other conditions, as there appeared to be less eye gaze and empathy correlations as judged by the positive relationship between Affective Empathy and participant eye fixation duration on mouth or nose area, as well as a negative correlation with eye area for the A_hV_lSAD condition. These findings are also relevant to H3 below.

H3: *State empathy will be higher for the highly realistic appearance and voice condition.* 5.3.3

The questionnaire results show that there was significant variance in empathy reports for both the realistic voice and realistic appearance, with both conditions leading to greater reports of state empathy on multiple metrics. As such, there is significant support for H3.

While the higher realism appearance provoked higher reports of both affective and compassionate empathy compared to lower appearance realism, there was a strong result for realistic voice on cognitive empathy reports, as well as the other two dimensions. There is an implication here that the real human voice offers affordances for emotion perception, a feature of cognitive empathy, in a way that is absent for our synthetic voice and character appearance stimuli. This implication needs further confirmation, but could mean that the inclusion of a human voice in emotional virtual character scenarios could be beneficial to the establishment of a perceptible character perspective.

Notably, the effects here differ from those of experiment 2C, which employed the same empathy questions in a study on variations of appearance realism alone. Compared to these results, the present work suggests that bimodal realism variations influence empathy to an extent that cannot be reached by variations in the unitary mode of appearance realism. This difference may be considered a distinction between perceptual content; without a voice, even a realistic character will seem less humanlike, regardless of emotion or appearance realism. In essence, the combination of voice and appearance together is itself an affordance for emotion perception, which can be influenced by very

subtle cues that are inextricable from real interactions, down to the inclusion of breaths in a vocal utterance [Jensen and Pedersen, 2016], which may explain our strong voice result. Nonetheless, it is this combination of voice and appearance realism which appears to provoke variation in reported empathy, in a finding which confirms our hypothesis and provides a direction for future work which seeks to examine how we perceive combinations of voice and appearance. [Stevenage and Neil, 2014].

H4a: *The highly realistic conditions will evoke lower uncanny ratings and higher ratings of appeal, realism and familiarity.* 5.3.3

H4b: *The mixed realism conditions will be considered more uncanny, with lower ratings of appeal, realism and familiarity.* 5.3.3

Finally, our uncanny valley hypotheses were also addressed by the analyses. Given the questionnaire results, there is evidence that the highly realistic condition was considered to be more appealing, and hence provide partial confirmation for H4a. This result is in line with experiment 2A using these particular characters, and highlight that certain combinations of high realism may surpass the uncanny valley. Despite previous mismatch approaches to measuring the uncanny valley yielding promising results [Mitchell et al., 2011b], our findings suggest that low levels of realism were independently considered more eerie than the highly realistic voice and appearance. One possible interpretation of this result is that the emotional dimension of this study, and resulting empathic responses, benefited from increased realism. It appears likely that the greater face detail and realism of the recorded human voice provide greater affordances for successful perception of emotions, and hence are less offputting than emotional expressions from less realistic characters.

Of note are our findings that trait Fantasising was positively correlated with zygomaticus activation in matched positive valence conditions (i.e. conditions where realism of voice and appearance were both high or both low). This finding could be interpreted as a form of emotional mimicry, specific to conditions of matched realism. The implication here is that people with certain empathic traits may be influenced by matched realism conditions to respond in a way that is comparable to human-human emotional interactions. In turn, mismatching realism (realistic voice and unrealistic appearance or vice versa) did not provoke this effect, which indicates that there could be an influence of affordance consistency on motor empathic responses from subjects more predisposed to fantasise. This finding could bear some significance to the literature on Empathy and the Uncanny Valley if it is confirmed by further studies.

A final note of interest, in response to a question posed by [MaCdorman et al., 2013],

is that there is no evidence in this work to suggest that the Uncanny Valley interferes with motor empathy. The emotion-congruent facial muscle activity that is recorded in our data suggests the occurrence of motor empathy that is not modulated by realism, or reports of character appeal or eeriness. Additionally, the fact that the lower realism characters were considered more eerie and less appealing would suggest that the realism mismatches were not necessarily considered uncanny. Overall, despite our finding on trait fantasisers, it is asserted here that motor empathy across emotions persisted despite reports of uncanniness.

5.5 Conclusion

The results of our experiment expand upon several lines of inquiry at the intersection of empathy and virtual characters.

Firstly, there appears to be a unique physiological response to realism and emotion in virtual characters that is demonstrated by this study; namely that the interaction between high realism and emotion was linked to increased zygomaticus major activity, with greatest activation recorded for the most realistic positive valence condition.

Participants' reports of state empathy indicate that realism influenced subjective perceptions of cognitive, affective and compassionate empathy. Appearance realism was mainly efficacious for affective and compassionate empathy measures, which is a useful finding for the construction of animated character-driven narratives that seek to evoke empathy. Additionally, given the advancement and widespread interaction with emotional text-to-speech systems, as exemplified by the emergence of GPT 4.0, there are questions to be asked about whether synthetic voices offer the full range of emotional affordances that are offered by a real voice. One of these affordances, it seems, is for cognitive empathy, as defined by our results on state empathy and the primacy of the real voice in this regard.

Our findings from the physiological data show that the Happy emotion provoked significantly higher Zyg muscle activity compared to Sad. We also found that Happy emotion led to increased visual attention to facial cues in these virtual character stimuli, in conjunction with congruent muscle activation. Conversely, there was an increase in Cor activity for Sad conditions. This result may be of benefit for research and development in games and educational training scenarios that require emotionally expressive virtual characters, given that a particular emotion expression may be more important for empathic narrative construction than the realism of the presented characters. Moreover, since there appears to be evidence for motor empathy in the EMG activity, this work

also contributes to the increasingly diverse literature on empathy, while the fundamental mechanisms behind it's occurrence remain to be fully addressed.

There is relevance in our results for future research into the Uncanny Valley, as this study highlights a unique physiological response to the interaction between trait empathy, emotion and realism; namely that zygomaticus responses to emotional expressions in virtual characters appear to differ for high trait fantasisers, in a manner which is consistent with realism mismatch theories of the uncanny valley. More work is needed to confirm, reject, or clarify this notion.

A significant limitation is that experiment 3 does not include a specific control condition, but explores the physiological effects while looking at videos of virtual characters in conditions of different combinations of factors such as realism. A possibility might be to include a third emotion, such as a neutral facial expression, as a potential control condition. In each video we included 5 seconds of static phase, where the virtual character is not animated. This static phase provided us with defined intervals, between which we could predict emotion congruent face muscle activity and concurrent eye fixation behaviour during dynamic emotion expressions. Nonetheless, an inclusion of a definite non-emotive or real human control condition would confirm the findings of this work.

Finally, given that this study was an exploration of empathy for emotionally valenced virtual human stimuli, similar confirmatory studies using neutral stimuli as well as other emotional dimensions would consolidate the findings of this work.

6 Conclusions

These combined works are proposed to contribute numerous findings on the nature of our interactions with virtual characters that are presented at greater and lesser degrees of human-likeness, in both human controlled and computer driven animated rigs. Our focus was on the combination of visual and vocal realism, and its effect on empathy and emotion perception. The following sections are devoted to explaining the contributions of this work, including the findings, as well as describing limitations for each experimental work that was conducted. This Chapter is also aimed at proposing future research directions that can be motivated from the bases of these findings and limitations.

6.1 Contributions

Given the broad scope of this work, there are a number of different domains in which the experiments outlined in Chapters 3-5 have provided novel contributions that may be used to guide future research in this area. These contributions may be broken down into several different dimensions, and are presented here separately for the sake of clarity.

Research Contributions

In reference to the final experiment, one of the major contributions of this work is in the recorded occurence of motor empathy for virtual characters. The EMG data suggest that emotion is afforded by characters at both high and low visual realism. However, self-reported empathy is measured to be experienced in gradations as a result of voice realism, where the presence of a voice evokes more empathy than its absence, and a real voice prompts more empathy than a synthetic one. As discussed in the background section and throughout this work, the combination of voice and appearance is best understood as an affordance for emotion and, by extension, empathy. A face without a voice or a voice without a face are fundamentally less human because they do not present the typical affordances that we experience in emotional social interactions with real people. Moreover, a real voice has been measured in this work to afford cognitive

empathy, the capacity to mentalise or take the perspective of another. This finding in particular should be a valuable contribution to empathy research that may guide some fruitful future works. This is shown by the results of the final experiment and the cross-experimental analysis. The question arises: why exactly does a real voice afford more cognitive empathy compared to a synthetic voice or a lone facial expression? The answer appears to be simple; we find it easier to understand an emotional expression from a human voice that is combined with an animated facial expression, and this in turn provides us with opportunities to take a character's perspective, in order to assess the relevance of their emotional state to further social actions, including joint attention. As outlined in the cross-experimental analysis, the presence of both real and synthetic voices influence reports of cognitive empathy compared to animated facial expressions alone. This finding hits upon something important about the nature of vocal utterance; it is a more direct outline of possible affordances compared to a facial expression. In the vocal utterances used by this thesis, the emotional state of the character is addressed directly. Emotion perception is easier with a vocal utterance, and hence understanding the perspective (i.e. experiencing cognitive empathy) is more easily achieved with the presence of a voice. A major contribution of this thesis stems from the the twin findings that a) realism in vocal emotional expressions can be understood as an affordance for cognitive empathy for virtual characters, compared to visual realism and b) visual realism, in terms of geometric and illumination realism, need not be considered necessary for developers to achieve cognitive empathy in comparison to voice realism.

Methodological Contributions

In addition, experiment 3 is also presented as a novel methodology for exploring and understanding empathy for virtual characters. There is a combination of synchronously recorded physiological measures which have been used to study virtual characters in the past, in tandem with the application of a state empathy questionnaire that had been developed for assessing empathy in human-human interactions. Future works can look to develop this methodology to study characters in different contexts and different conditions of realism and emotion. Moreover, since we appeared to observe an influence of realism on physiological data, with the most realistic happy condition provoking greater Zygomaticus major activity, this methodology may be employed in comparing facial responses to human emotions and virtual humans at other degrees of realism. This novel methodology may be applied in research and development in the use of embodied agents, voice agents, video game characters and other relevant applications of animated faces and voices.

Similarly, experiments 1A and 1B were conducted using avatars for video conferencing,

which had previously been an unstudied niche in virtual character interaction. Given that these experiments were conducted in a period of repeated pandemic related quarantine policies, this methodology was particularly relevant to the time period and may well again be in the future. The interview format for emotion elicitation measurement and overall virtual character perception work may be considered a second novel methodology for studies which seek to understand how emotion communication can be achieved with 3D avatars.

Subjectivity Contributions

Overall, the results from our experiments suggest that emotion perception and elicitation can be achieved with virtual characters, with faces and voices at varying level of realism, in a fashion which can be modulated by subjective traits. For example, in the final experiment, there was greater activation of the Zygomaticus major muscle in response to virtual characters with matched levels of realism (i.e. low realism voice and appearance conditions, as well as high realism voice and appearance conditions, when presented together), specific to users who scored highly on trait fantasising, which is considered one of the elements of empathy as measured by the interpersonal reactivity index. This novel finding supports the notion that our personal subjective traits shape our responses to virtual characters, in a manner which is observable at the physiological level. Other traits were observed to have notable effects, particularly in our avatar interaction work. While we used these findings to motivate further work on empathy for independent virtual characters, there are standalone findings that will be useful for future social-psychological research with humanoid avatars. Namely, the Big Five personality factor of Agreeableness was measured to positively influence the Sense of Embodiment in humanoid avatars, in both video-conferencing and immersive 3D settings. This novel finding is at odds with previous similar works, and so must be interpreted cautiously, and confirmed by further studies. Nonetheless, there is an implication here that a tendency to respond agreeably to novel scenarios and stimuli leads some individuals to report greater senses of embodiment in a given avatar. On a related note, there may be a tendency for people high in trait agreeableness to succumb to demand characteristics of experimental settings, in that they are more likely to want to 'answer correctly' in some form that they interpret from the design of an experiment. This limitation is expanded upon later in this section. These findings are applicable to research work which seeks to understand the subjectivity of experiences with virtual characters, and may prove useful for personalised educational simulations that involve such characters.

Theoretical Contributions

This work contributes significantly to understanding the Uncanny Valley phenomenon. In Chapters 4 and 5, significant coverage is given to understanding the nature of the Uncanny. In this work, the appearance of the characters appeared to benefit from improved visual realism, in terms of being considered less eerie, more appealing and generating higher reports of empathy from participants. As such, our findings suggest that higher levels of visual realism that are achievable with modern commercial processors and advanced animation techniques are not considered uncanny compared to less realistic counterparts. This suggests a future where the valley may be traversed, with specific regards to vocal-visual realism in virtual characters. It is important to note that this may not apply to robots, or more advanced forms of realism such as behavioural and motion realism, as they were not the focus of this thesis. Indeed, behavioural realism for virtual humans and robots may be considered the most difficult to master. While our work suggests that vocal and visual realism may improve perceptions of virtua l characters, there remains an open question as to whether they will ever be able to act human in a convincing enough manner as to avoid the Uncanny Valley. Nevertheless, there is some evidence, linked to our trait Fantasising finding, that matched modes of realism may ease emotion perception and empathy, compared to characters that are presented with realistic voices and less realistic faces or vice versa. This design feature may serve as a recommendation for developers that aim to use emotionality to guide users through narratives in film and games.

Material Contributions

The stimuli created for Chapters 4 and 5 are open to others who seek to measure empathy in virtual characters. This includes real-time captured motion for generating the movements of the characters in Chapter 4, as well as offline animated facial and head movement that was used for Chapter 5. Moreover, this worked employed voices that express vocal emotion, captured from real actors as well as generated synthetically. These vocal utterances were composed from a dictionary of affective words and phrases, and their novel composition has been seen in this work to display positive and negative valence emotions. Moreover, the significant amount of survey and physiological data generated by our studies may be used in meta-reviews, replication studies and other related scientific works.

Design Contributions

Taken together, our findings provide guidelines that can be used in the design development of virtual characters for numerous purposes, including for use in games, television and immersive media, as well as avatar use in medical scenarios, such as for brain damaged patients who are now able to use avatars for the purpose of general communication.

- **Visual Realism:** Since visual realism did not detract from emotion perception of perceived appeal of the characters, visual realism may be considered a positive choice when combined with realistic voices. However, this may not apply to certain forms of interaction, for example pedagogical settings, if there follow other studies which determine appeal and empathy to distract from student engagement with learning content.
- **Vocal Realism:** Similar claims may be made for designers who create characters with realistic voices. Particularly, our analyses reveal an overall preference for real human voices in virtual characters. While voice synthesis for non-embodied agents (e.g., Alexa) may prove to be more cost effective, when it comes to simulating human likenesses, virtual character design does appear to benefit from real human voices in scenarios that involve emotional expressions.
- **Empathy & Emotion:** These factors of realism are specific to characters expressing emotions. While future work, discussed below, should aim to understand how generalizable the findings on realism could be to other scenarios, there appears to be both an empirical and an intuitive relationship between realism, empathy and emotion. Designers who seek to maximise both intensity of emotion and clearly outline a particular valence of emotion can be encouraged by this work to use realism to improve these aspects in narrative scenarios as well as contexts where patients are aiming to express emotions through the use of avatars for communication purposes.

6.2 Limitations and Future Work

Following these experimental works, emotion recognition between humans and virtual characters is considered to be a domain ripe for further research, both through analysing social dynamics, developing new technical methods or applying knowledge of perceptual processes of emotion. It was through this research that we began to formulate ideas that could be valuable for understanding human emotion in general, but more specifically the means by which humans are able to recognise and respond to them in virtual characters at different degrees of realism.

In view of the evidence in the present thesis, subsequent research should seek to consolidate our results with investigations into visual effects on valence outcomes. Work from Wisessing et al. [Wisessing et al., 2020] has suggested that positive and negative affect conditions can be enhanced by bright and dark key-light illumination respectively. Lighting conditions studied in conjunction with manipulated avatar voice and appearance could expand on these results, and could have wide-reaching implications for any research or clinical setting that would benefit from mood enhancement, such as video conferences or clinical treatments.

While experiment 1C focused on visuo-motor synchrony, there is important work to be done on visuo-tactile perception and face-ownership, considering behavioural manipulations such as threat induction have already proved fruitful on this topic in body-ownership research [Ma and Hommel, 2013]. Similar future studies that develop greater real time animation detail, such as those in deep fakes, could also investigate real-time facial rigs with a greater depth of expressive features, as the software in experiments 1A and 1B involve the standard 51 expressions (i.e. Apple's ARKit face tracking which is the current standard for real-time). Other future work to be considered as implications of this study would be a more natural conversational structure between participants who are known or anonymous to each other. This work could shed a new light on interactional processes in avatar use.

There are constraints on the conclusions we can draw from the data, given the sample sizes and potential for confounds inherent in the design of non-laboratory experiments (as per experiments 1A and 1B, as well as 2A, 2B and 2C).

Both experiments 1A and 1B used cartoon avatars, but the main experiment used more realistic personalized digital characters. This may have introduced elements of the Uncanny Valley effect, although there is no evidence for this in the negative valence results, this also as not measured for these experiments, and was only considered to be an influential factor from experiment 2A onwards.

There is sufficient justification outlined here for further investigations into the interplay between personal characteristics and the emergence of subjectively non-veridical perceptions in both 3D or screen-mediated environments that involve digital characters with varying levels of realistic detail.In terms of experiment 1C, which investigated subjectivity and its effects on embodiment, there is sufficient evidence to suggest that trait influences can be found in avatar embodiment scenarios. On the back of prior work, these findings suggest that trait assessments are important to contextualise future studies on avatar embodiment. Of course, in order to confirm this notion analyses will need to be conducted on much larger samples, with greater variation in embodiment conditions.

In studies 2A, 2B, 2C and 3, our low level of realism was created by lowering the facial rig complexity, and the geometric and texture complexity to mimic characters currently seen across many applications, while retaining the same identity, motion, voice, lighting and rendering conditions as the high fidelity version. While eeriness ratings were low for our high fidelity characters, it is possible that ratings might reach even lower levels with higher quality facial animation. We used real-time facial tracking and solving with only minimal clean-up in this study, and previous work has shown that lower quality facial animation can appear more eerie on realistic characters [McDonnell, 2012]. Future work should investigate the effect of facial tracking and solving.

Future work should also address different levels of voice realism in dialogue with a virtual character, as opposed to just constrained video-question formats. It would also be a useful direction for addressing the domain of empathy for virtual characters in greater detail. A limitation on this thesis may be the constraints placed upon the interactions with the characters. Real-time lip sync and gesture generation animations in tandem with high functioning dialogue systems provide an interesting future avenue for this kind of perceptual research.

In human-avatar interaction research, perceptually salient modulations of voice, such as the ones described in this research, have previously been proposed to influence ratings of attractiveness and immersion and task performance [Kao et al., 2021] for gaming environments. Emotionally valenced game-based scenarios between human controlled avatars and agents could be a fruitful future direction for research.

Researchers should seek to replicate this kind of work using TTS voices in 3D VR contexts. This work would expand upon previous findings in experiments that have analysed perceptions of feedback from virtual agents in 3D VR. Such work has suggested that auditory feedback is preferable above other forms, namely behavioural and visual feedback, when partnered with virtual agents in a task-based environment [Baxter et al., 2021], which could serve to elucidate to the effect of speech feedback from VR agents with both natural, morphed or clearly synthetic voices, in both task-based and emotionally valenced scenarios such as ours.

Similarly, experimental work might seek to employ lower levels of realism when aligned with multiple levels of voice realism. This may help to process perceptual mismatch data for the purpose of visualisation in terms of the uncanny valley.

It may also be argued that for ethical purposes, an obviously synthetic voice could be advantageous in gaining agent rapport, to retain trust with the user who may in the future find it difficult to determine if a virtual character is real, based on its appearance.

In terms of methodological constraints, future work may seek to measure the phenomenon of expectation. For example, isolated exposures during between-subjects studies can be considered experientially different to those conducted within-subjects. It is possible that expectation effects are potentially influential in within subjects designs for measuring the uncanny valley, which would be a significant limitation for any within-subjects experiments [Lush et al., 2020]. Moreover, recent research that has sought to cast doubt on previous results from manipulating visuo-tactile information to demonstrate the 'Rubber Hand Illusion', a famous SoE induction experiment [IJsselsteijn et al., 2006], are based on a proposed failure for its control of demand characteristics. New knowledge on this topic implicates the role of individual traits on the outcome of results in studies that rely on perceptual senses of embodiment [Lush, 2020]. Future research should directly compare between-subjects and within-subjects designs for these kinds of studies, in an effort to understand the influence of demand characteristics.

While the mismatch of vocal and visual modes of realism may play a role in perceived uncanniness [Mitchell et al., 2011a] which may also explain this deviation in results and similar work has claimed that combinations of multi-modal realism conditions produce uncanny effects [Macdorman et al., 2008, Burleigh et al., 2013] more reliably than the single variable of visual realism, these findings were not observed in our experiments. It is necessary to caveat that the experiments from Chapters 4 and 5 have not achieved clear cut and indisputable results dictating that highly mismatched appearance and voice realism influence social presence or induce the uncanny valley effect. Further studies should incorporate differing levels of appearance realism employed in conjunction with different voices in order to achieve further clarity in this domain. Additional studies including multi-modal variations in realism must be pursued to clarify this relationship. Future work could also aim to replicate these same experiments in a virtual reality context, with an increasing role to be played by immersive environments on empathic responses [Slater et al., 2018, van Loon et al., 2018b]. Further studies which seek to find alternative markers for empathic responses to realistic virtual humans will also add value to the work described in this paper.

Several limitations exist that are specific to the identities of the virtual characters used. Firstly, in order to reduce possible confounds, each of the virtual humans experiments (2A, 2B and 3) used multiple ethnicities and two genders of character. The aim was to include a diverse range of characters so that participant gender or ethnicity would not be a factor in their responses to each block. This move was precautionary, as previous results from experiment 2A have suggested that these four particular characters are not perceived differently. However, given that similar human-human studies have found differences in emotional mimicry responses using our same EMG measurements

[Shenli Peng and Hu, 2021], a further development from this study should use a more heterogenous participant sample in order to fully assess any effects of 'in-group' virtual character perceptions. This problem is echoed in the Uncanny Valley literature [Saneyoshi et al., 2022]. As such, there is a limitation to the other experiments, 1A, B and C, as well as 2B, that there may have been a difference in results if multiple characters had been used.

An additional suggestion for future work would be to include further dimensions of realism. While quite narrowly constrained in our case, future studies of empathy for virtual characters could include real human stimuli, extremely unrealistic point light avatars, and other forms of realism to do with body movement or social behaviour. In this regard, our results for the interaction effect of realism and emotion on physiological responses are limited to the constrained range of realism options employed in this study.

In general, this thesis as a whole can be seen to contain several limitations. One of the core progressions of this work is in the change from focus on avatar perception to character model perception. While the virtual human models used for the realism experiments in Chapters 4 and 5 may be used as avatars, and hence the findings may be considered generalizable, it is important for future works to address this distinction when it comes to perception of the Uncanny Valley and empathic responses. Comparative studies that use pregenerated animations and real time human-controlled responses can be seen as a core future focus in this domain. Similarly, users who are embodied as avatars may also be seen as likely to perceive empathy and character appeal differently, and hence this may also be a useful future direction for further research. Furthermore, the specified emotion and realism dimensions employed in this work were not deployed in avatar-avatar interactions, nor were they conducted in immersive 3D virtual environments. As such, studies that differentiate between these aspects of study design will be valuable contributions to the field.

Another limitation may stem from our measurements of subjective traits. While the final studies used the Interpersonal Reactivity Index to assess trait empathy, as justified by the literature, there are also numerous other tools of varying levels of depth and robustness that may also be useful in this domain. Our narrow application of one trait empathy measurement may be considered a limitation. However, given the novel application of this tool to measure trait empathisers relationships to realism and emotional expressions in virtual characters, future works using other trait measurement tools would be useful to further explore this niche. Comparable limitations and future works may also be highlighted by our use of the Five Factor personality assessment and trait body awareness models.

Finally the most significant limitation, as mentioned for each particular experiment above, comes from the dimensions of realism and emotion applied across the studies. While the contributions of this thesis stand alone in regards to vocal realism and visual realism together with positive and negative valence emotions, there is a need to contextualise these findings with further works. Such works should assess different combinations and ranges of our realism factors, as well as interactive, behavioural and animation realism. Similarly, emotions beyond simply portrayed happy and sad emotions, as well as neutral stimuli and emotions at varying degrees of intensity, would be beneficial to the study of empathy for virtual characters.

6.3 Conclusions

Despite these limitations, this thesis is proposed to advance the study of virtual characters, specifically those that resemble humans. On these terms, this work has shown that virtual characters at high levels of vocal and visual realism can be viewed as transcending the Uncanny Valley. However, this conclusion must be viewed in context; the idea that a fully functioning virtual human, complete with advanced behavioural realism and interactive responsivity, has surpassed the Uncanny Valley is beyond the scope of this work. Speculatively, we may say that it is unlikely that an interactive virtual human at high realism will be able to completely shed the sense of the Uncanny, given that it is not yet truly possible for a virtual human to behave in precisely the same fashion as a real human. Systems for the generation of deep fakes will become more relevant in this domain, given their intention to fully replicate human-likenesses, but the generation of 3D characters that are realistic depictions (rather than replications) of real people can be viewed as fundamentally separate to this form of realism. However we may expect that deep fake realism will be studied in the context of the Uncanny Valley.

As highlighted, the results of our experimental work are best viewed under the theoretical constraints of affordances and direct perception; in this light, we may frame the experiences of empathy and physiological responses to emotional expressions as responses to perceptual cues (affordances) that invite reactions from the user (mentalising, affective contagion, sympathy, motor empathy). While in our specific lab-based studies with virtual characters there was not further interactivity invited from the user, in general these empathic reactions allow for further social actions to take place (for example, to comfort someone sad, to attend to the same object as someone who is scared and so forth). Hence, the results may be described as relevant to the aforementioned Computers as Social Actors paradigm. Moreover, to couch these results in the language of affordances allows us to describe the treatment of humanlike objects as both social actors and ob-

jects. These are after all only models designed to resemble humans, rather than any kind of extremely lifelike artifical human that may trick a user into seeing it as truly real. Hence, despite the direct perception of emotion and the corresponding empathic reactions, and despite the fact that these virtual characters are able to afford humanlike emotion and corresponding social perceptual mechanisms, both in voice and appearance, they are still objects that exist in a virtual space. As such, the notion that objects in a virtual space can be so influential over our socio-affective responses is best understood as their affording social action, through the combination of vocal and visual emotional expressions.

In the background section, a significant effort was made to unify the focus of these research studies with contemporary advancements in cognitive science, specifically to do with the approach to affordances that is taken by the interface theory of perception, the constructivist theory of emotion, and the direct perception of empathy. Taken together, these theories suggest an active role of the perceiver in the construction of both realism and emotion, in a manner which prioritises experience over cognitive inference. Our final findings are specifically relevant here. Firstly, while both vocal and visual realism were measured to influence self-reported empathic responses, they did not influence motoric responses to emotion. This suggests a role of facial mimicry in the construction of an emotion concept, while simultaneously suggesting that the top down cognitive mediation required to report empathy does not play a role in the physiological response to an emotional expression in a virtual character. Hence the direct perception of the emotional affordances, as measured in the facial responses, may be understood as a separate kind of empathy to the self reports that may be mediated by cognitive inferences.

Secondly, the specific trait of fantasising, which measures the capacity to empathise with external, non-human actors in the world, was measured to influence physiological responses to positively valenced emotions at both high and low levels of realism. This work suggests a place among these contemporary views of cognitive science; the subject observes cues in the face and voice of the virtual character, cues which are themselves affordances for acting in the world. These cues are married to previously held concepts, dependent on the subject and their own capacities for empathy. These capacities and concepts are thus measurable in trait fantasisers' facial reactions to displays of emotion at matched levels of realism, where the consistency of the cues provoked stronger physiological reactions. This unique finding and its unique place in the literature is proposed to sit in between these theories; faces and voices are modes of the perceptual interface that are particularly salient to our social interactions, and the consistency of their presentation enables the construction of emotion which is highlighted by the strength of the facial response in the subject.

Overall, vocal and visual realism, specific to our definitions, were not as influential on empathic responses as expected in our final study. This is likely due to the range of affordances offered by the stimuli. The virtual faces, within our defined scope of realism, offer the same facial expressions, the same facial features and the same movements of the mouth, eyes and brow. The difference in realism was due to the graphical detail; there was greater geometric complexity and illumination quality in the more realistic facial representations, and this detail was considered likely to influence physiological responses. However, the difference was only reflected in self-reports of affective and compassionate empathy. Hence we may conclude that the affordances for both conditions of visual realism are similar enough to provoke similar levels of cognitive empathy, in that the emotion could be recognised and understood, and similar levels of motor empathy, where the direct experience of the emotional content provoked the same reactions in the zygomaticus and corrugator muscles. Additionally, the experiential content of the emotion expression was influential over eye fixation behaviour where visual and vocal realism were not. This conclusion is important for the field of computer graphics, in the sense that developers need not focus on geometric or illumination realism in order to afford cognitive empathy.

In sum, this thesis may conclude with evidence to support the Computers As Social Actors paradigm, given the self reports of empathy across virtual character conditions, as well as face muscle responses in the final study comparable to the same face muscle responses to emotions in real human interactions. However, this finding would need further confirmation in interactive settings, to test whether conversational dynamics may be influenced by realism and emotion in similar settings. Moreover, the results of our experiments confirm that there is a role for individual traits in shaping the perception of virtual characters, whether they come in the form of embodied avatars or individual virtual agents. On the final view, these findings provide answers to our main research questions, namely that vocal realism tends to influence cognitive empathy, while visual realism does not. Reports also indicated that both affective and compassionate empathy are influenced by both visual and vocal realism. Finally, matched modes of realism tended to evoke greater motor empathy to users who were more likely to exhibit traits involved in empathic fantasising, which provides an interesting conclusion to our secondary research question. These findings, as outlined above, additionally provide useful future directions for research which seeks to address how realism and emotion influence our empathic responses to different kinds of virtual characters in different kinds of scenarios.

Bibliography

[Abdulrahman and Richards, 2022] Abdulrahman, A. and Richards, D. (2022). Is natural necessary? human voice versus synthetic voice for intelligent virtual agents. *Multimodal Technologies and Interaction*, 6.

[Ahn et al., 2011] Ahn, S. J.-G., Fox, J., and Bailenson, J. (2011). Avatars. pages 695–702.

[Araujo et al., 2021] Araujo, V., Dalmoro, B., and Musse, S. R. (2021). Analysis of charisma, comfort and realism in cg characters from a gender perspective. *Visual Computer*, 37:2685–2698.

[Argelaguet et al., 2016] Argelaguet, F., Hoyet, L., Trico, M., and Lécuyer, A. (2016). The role of interaction in virtual embodiment: Effects of the virtual hand representation. volume 2016-July, pages 3–10. IEEE Computer Society.

[Auyeung et al., 2015] Auyeung, B., Lombardo, M. V., Heinrichs, M., Chakrabarti, B., Sule, A., Deakin, J. B., Bethlehem, R. A., Dickens, L., Mooney, N., Sipple, J. A., Thiemann, P., and Baron-Cohen, S. (2015). Oxytocin increases eye contact during a real-time, naturalistic social interaction in males with and without autism. *Translational Psychiatry*, 5.

[Aylett and Ballin, 2013] Aylett, R. and Ballin, D. (2013). *Intelligent virtual agents*. Springer.

[Badler, 1997] Badler, N. I. (1997). Real-time virtual humans real-time virtual humans real-time virtual humans. *Copyright*, pages 28–36.

[Bailenson and Blascovich, 2004] Bailenson, J. N. and Blascovich, J. (2004). Avatars. In *Encyclopedia of Human-Computer Interaction, Berkshire Publishing Group*, pages 64–68. Berkshire Publishing Group.

[Bailenson et al., 2003] Bailenson, J. N., Blascovich, J., Beall, A. C., and Loomis, J. M. (2003). Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin*, 29(7):819–833.

[Bailenson et al., 2006] Bailenson, J. N., Yee, N., Merget, D., and Schroeder, R. (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. 15:359–372.

[Bailenson et al., 2016] Bailenson, S. Y. O. . J., Krämer, N., and Li, B. (2016). Let the avatar brighten your smile: Effects of enhancing facial expressions in virtual environments. *PLoS ONE*, 11(9):e0161794.

[Bailey et al., 2016] Bailey, J. O., Bailenson, J. N., and Casasanto, D. (2016). When does virtual embodiment change our minds? *Presence: Teleoperators and Virtual Environments*, 25:222–233.

[Baird et al., 2018] Baird, A., Parada-Cabaleiro, E., Hantke, S., Burkhardt, F., Cummins, N., and Schuller, B. (2018). The Perception and Analysis of the Likeability and Human Likeness of Synthesized Speech. In *Proc. Interspeech 2018*, pages 2863–2867.

[Balconi and Canavesio, 2016] Balconi, M. and Canavesio, Y. (2016). Is empathy necessary to comprehend the emotional faces? the empathic effect on attentional mechanisms (eye movements), cortical correlates (n200 event-related potentials) and facial behaviour (electromyography) in face processing. *Cognition and Emotion*, 30:210–224.

[Bargh and Chartrand, 2000] Bargh, J. A. and Chartrand, T. L. (2000). The mind in the middle: A practical guide to priming and automaticity research.

[Barrett, 2011] Barrett, L. F. (2011). Constructing emotion. *Psychological Topics*, 20:359–380.

[Barrett, 2013] Barrett, L. F. (2013). Psychological construction: The darwinian approach to the science of emotion. *Emotion Review*, 5:379–389.

[Barrett, 2016] Barrett, L. F. (2016). *Navigating the Science of Emotion*, pages 31–63. Elsevier Inc.

[Barrett, 2017] Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Pan Macmillan.

[Barrett et al., 2019] Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., and Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring

emotion from human facial movements. *Psychological science in the public interest : a journal of the American Psychological Society*, 20:1–68.

[Bartneck et al., 2007] Bartneck, C., Kanda, T., Ishiguro, H., and Hagita, N. (2007). Is the uncanny valley an uncanny cliff? pages 368–373.

[Bavelas et al., 1986] Bavelas, J. B., Black, A., Lemery, C. R., and Mullett, J. (1986). " i show how you feel": Motor mimicry as a communicative act. *Journal of personality and social psychology*, 50(2):322.

[Baxter et al., 2021] Baxter, M., Bleakley, A., Edwards, J., Clark, L., Cowan, B. R., and Williamson, J. R. (2021). You, move there!: Investigating the impact of feedback on voice control in virtual environments. Association for Computing Machinery.

[Baños et al., 2004] Baños, R. M., Botella, C., Alcañiz, M., Liaño, V., Guerrero, B., and Rey, B. (2004). Immersion and emotion: Their impact on the sense of presence. *CYBERPSYCHOLOGY and BEHAVIOR*, 7.

[Beck, 2008] Beck, A. (2008). Realistic simulation of emotion by animated characters. *PhD diss., Portsmouth University, ND*.

[Beer et al., 2015] Beer, J. M., Smarr, C.-A., Fisk, A. D., and Rogers, W. A. (2015). Younger and older users recognition of virtual agent facial expressions. *International Journal of Human-Computer Studies*, 75:1–20.

[Belin et al., 2011] Belin, P., Bestelmeyer, P. E., Latinus, M., and Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*, 102(4):711–725.

[Ben-David et al., 2011] Ben-David, B. M., Lieshout, P. H. V., and Leszcz, T. (2011). A resource of validated affective and neutral sentences to assess identification of emotion in spoken language after a brain injury. *Brain Injury*, 25:206–220.

[Bendel, 2019] Bendel, O. (2019). The synthetization of human voices. *Ai & Society*, 34:83–89.

[Bergson et al., 2018] Bergson, H., Brereton, C. S. H., and Rothwell, F. (1900/2018). *Laughter: An Essay on the Meaning of the Comic*. Franklin Classics.

[Bevacqua et al., 2012] Bevacqua, E., De Sevin, E., Hyniewska, S. J., and Pelachaud, C. (2012). A listener model: introducing personality traits. *Journal on Multimodal User Interfaces*, 6(1-2):27–38.

[Biocca, 1997] Biocca, F. (1997). The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments [1]. *Journal of Computer-Mediated Communication*, 3(2). JCMC324.

[Biocca et al., 2001] Biocca, F., Burgoon, J. K., and Stoner, G. M. (2001). Criteria and scope conditions for a theory and measure of social presence see profile.

[Blow et al., 2006] Blow, M., Dautenhahn, K., Appleby, A., Nehaniv, C. L., and Lee, D. C. (2006). Perception of robot smiles and dimensions for human-robot interaction design.

[Bogotá and Artese, 2022] Bogotá, J. D. and Artese, G. F. (2022). A husserlian approach to affectivity and temporality in affordance perception. In Bogotá, J. D. and Artese, G. F., editors, *Affordances in Everyday Life. A Multidisciplinary Collection of Essays*, pages 181–190. Springer.

[Bouchard et al., 2013] Bouchard, S., Bernier, F., Boivin, E., Dumoulin, S., Laforest, M., Guitard, T., Robillard, G., Monthuy-Blanc, J., and Renaud, P. (2013). Empathy toward virtual humans depicting a known or unknown person expressing pain. *Cyberpsychology, Behavior, and Social Networking*, 16(1):61–71.

[Burden and Savin-Baden, 2019] Burden, D. and Savin-Baden, M. (2019). *Virtual Humans: Today and Tomorrow*. Chapman & Hall/CRC Artificial Intelligence and Robotics Series. CRC Press.

[Burin et al., 2019] Burin, D., Pignolo, C., Ales, F., Giromini, L., Pyasik, M., Ghirardello, D., Zennaro, A., Angilletta, M., Castellino, L., and Pia, L. (2019). Relationships between personality features and the rubber hand illusion: an exploratory study. *Frontiers in psychology*, 10:2762.

[Burleigh et al., 2013] Burleigh, T. J., Schoenherr, J. R., and Lacroix, G. L. (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(3):759–771.

[Cabral et al., 2017] Cabral, J. P., Cowan, B. R., Zibrek, K., and McDonnell, R. (2017). The Influence of Synthetic Voice on the Evaluation of a Virtual Character. In *Proc. Interspeech 2017*, pages 229–233.

[Carney, 2020] Carney, J. (2020). Thinking avant la lettre: A review of 4e cognition. *Evolutionary studies in imaginative culture*, 4(1):77–90.

[Carra et al., 2019] Carra, E., Santoni, C., and Pellacini, F. (2019). Grammar-based procedural animations for motion graphics. *Computers & Graphics*, 78:97–107.

[Cebolla et al., 2019] Cebolla, A., Herrero, R., Ventura, S., Miragall, M., Bellosta-Batalla, M., Llorens, R., and Baños, R. M. (2019). Putting oneself in the body of others: A pilot study on the efficacy of an embodied virtual reality system to generate self-compassion. *Frontiers in Psychology*, 10:1521.

[Cebolla et al., 2016] Cebolla, A., Miragall, M., Palomo, P., Llorens, R., Soler, J., Demarzo, M., García-Campayo, J., and Baños, R. M. (2016). Embodiment and body awareness in meditators. *Mindfulness*, 7:1297–1305.

[Cerda et al., 2021] Cerda, L., Fauvarque, A., Graziani, P., and Del-Monte, J. (2021). Contextual priming to increase the sense of presence in virtual reality: exploratory study. *Virtual Reality*, 25:1105–1112.

[Chalmers, 2017] Chalmers, D. J. (2017). The virtual and the real. *Disputatio*, 9:309–352.

[Chartrand and Bargh, 1999] Chartrand, T. L. and Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction.

[Cheetham et al., 2009] Cheetham, M., Pedroni, A., Antley, A., Slater, M., and Jäncke, L. (2009). Virtual milgram: empathic concern or personal distress? evidence from functional mri and dispositional measures. *Frontiers in human neuroscience*, 3:802.

[Cheetham et al., 2015] Cheetham, M., Wu, L., Pauli, P., and Jancke, L. (2015). Arousal, valence, and the uncanny valley: Psychophysiological and self-report findings. *Frontiers in psychology*, 6:126882.

[Chella et al., 2018] Chella, A., Belpaeme, T., Holland, O. E., Linson, A., Clark, A., Ramamoorthy, S., and Friston, K. (2018). the active inference approach to ecological perception: General information dynamics for natural and artificial embodied cognition. 5.

[Chen et al., 2005] Chen, B.-Y., Ono, Y., and Nishita, T. (2005). Character animation creation using hand-drawn sketches. *The Visual Computer*, 21:551–558.

[Cheymol et al., 2023] Cheymol, A., Fribourg, R., Lécuyer, A., Normand, J.-M., and Argelaguet, F. (2023). Beyond my Real Body: Characterization, Impacts, Applications and Perspectives of "Dissimilar" Avatars in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–12.

[Ciechanowski et al., 2019] Ciechanowski, L., Przegalinska, A., Magnuski, M., and Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*, 92:539–548.

[Cobb-Clark and Schurer, 2012] Cobb-Clark, D. A. and Schurer, S. (2012). The stability of big-five personality traits. *Economics Letters*, 115(1):11–15.

[Cohn et al., 2007] Cohn, J. F., Ambadar, Z., and Ekman, P. (2007). Observer-based measurement of facial expression with the facial action coding system. *The handbook of emotion elicitation and assessment*, 1(3):203–221.

[Costa et al., 2013] Costa, M. R., Kim, S. Y., and Biocca, F. (2013). Embodiment and embodied cognition. volume 8021 LNCS, pages 333–342. Springer Verlag.

[Cowan et al., 2014] Cowan, D. G., Vanman, E. J., and Nielsen, M. (2014). Motivated empathy: The mechanics of the empathic gaze. *Cognition and Emotion*, 28:1522–1530.

[Crescentini et al., 2016] Crescentini, C., Chittaro, L., Capurso, V., Sioni, R., and Fabbro, F. (2016). Psychological and physiological responses to stressful situations in immersive virtual reality: Differences between users who practice mindfulness meditation and controls. *Computers in Human Behavior*, 59:304–316.

[Cross et al., 2015] Cross, E. S., Ramsey, R., Liepelt, R., Prinz, W., and Hamilton, A. F. d. C. (2015). The shaping of social perception by stimulus and knowledge cues to human animacy. 371(1686).

[Crumpton and Bethel, 2016] Crumpton, J. and Bethel, C. L. (2016). A survey of using vocal prosody to convey emotion in robot speech. *International Journal of Social Robotics*, 8:271–285.

[C.S. Oh and Welch, 2018] C.S. Oh, J. N. B. and Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers Robotics AI*, 5.

[Cummins, 2017] Cummins, R. G. (2017). Eye tracking. *The International Encyclopedia of Communication Research Methods (1-9). Texas Tech University, USA. https://doi.org/10.1002/9781118901731. iecrm0099*.

[da Costa et al., 2021] da Costa, N. M., Bicho, E., Ferreira, F., Vilhena, E., and Dias, N. S. (2021). A multivariate randomized controlled experiment about the effects of mindfulness priming on eeg neurofeedback self-regulation serious games. *Applied Sciences (Switzerland)*, 11.

[Davis, 1980] Davis, M. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, 10:85–103.

[Davis, 1983] Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, 44(1):113.

[Davis et al., 1980] Davis, M. H., Davis, M. P., Davis, M., Davis, M., Davis, M., Davis, M., Davis, M., Davis, F. C., Davis, H. A., and Davis, I. W. (1980). A multidimensional approach to individual differences in empathy.

[de Borst and de Gelder, 2015] de Borst, A. W. and de Gelder, B. (2015). Is it the real deal? perception of virtual characters versus humans: An affective cognitive neuroscience perspective. *Frontiers in Psychology*, 6.

[de Hamilton, 2016] de Hamilton, A. F. (2016). Gazing at me: The importance of social meaning in understanding direct-gaze cues. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371.

[de Kervenoael et al., 2020] de Kervenoael, R., Hasan, R., Schwob, A., and Goh, E. (2020). Leveraging human-robot interaction in hospitality services: Incorporating the role of perceived value, empathy, and information sharing into visitors' intentions to use social robots. *Tourism Management*, 78.

[De Luca, 2003] De Luca, G. (2003). Fundamental concepts in emg signal acquisition. *Copyright Delsys Inc*.

[de Melo et al., 2013] de Melo, C., Carnevale, P., and Gratch, J. (2013). People's biased decisions to trust and cooperate with agents that express emotions. In *Proceedings from International Conference on Autonomous Agents and Multiagent Systems*.

[de Melo et al., 2014] de Melo, C. M., Gratch, J., and Carnevale, P. J. (2014). Humans versus computers: Impact of emotion expressions on people's decision making. *IEEE Transactions on Affective Computing*, 6(2):127–136.

[de Melo et al., 2023] de Melo, C. M., Gratch, J., Marsella, S., and Pelachaud, C. (2023). Social functions of machine emotional expressions. *Proceedings of the IEEE*, pages 1–16.

[de Rooij et al., 2017] de Rooij, A., van der Land, S., and van Erp, S. (2017). The creative proteus effect: How self-similarity, embodiment, and priming of creative stereotypes with avatars influences creative ideation. In *Proceedings of the 2017 ACM*

SIGCHI Conference on Creativity and Cognition, C&C '17, page 232–236, New York, NY, USA. Association for Computing Machinery.

[De Vignemont and Singer, 2006] De Vignemont, F. and Singer, T. (2006). The empathic brain: how, when and why? *Trends in cognitive sciences*, 10(10):435–441.

[Debruille et al., 2012] Debruille, J. B., Brodeur, M. B., and Porras, C. F. (2012). N300 and social affordances: A study with a real person and a dummy as stimuli. *PLoS ONE*, 7.

[Devesse et al., 2018] Devesse, A., Dudek, A., van Wieringen, A., and Wouters, J. (2018). Speech intelligibility of virtual humans. *International Journal of Audiology*, 57:908–916.

[Dewez et al., 2019a] Dewez, D., Fribourg, R., Sanz, F. A., Hoyet, L., Mestre, D., Slater, M., Lécuyer, A., and Argelaguet, F. (2019a). Influence of personality traits and body awareness on the sense of embodiment in virtual reality.

[Dewez et al., 2019b] Dewez, D., Fribourg, R., Sanz, F. A., Hoyet, L., Mestre, D., Slater, M., Lécuyer, A., and Argelaguet, F. (2019b). Influence of personality traits and body awareness on the sense of embodiment in virtual reality. pages 1–12.

[Diel et al., 2021] Diel, A., Weigelt, S., and Macdorman, K. F. (2021). A meta-analysis of the uncanny valley's independent and dependent variables. *J. Hum.-Robot Interact.*, 11(1).

[Diel et al., 2022] Diel, A., Weigelt, S., and Macdorman, K. F. (2022). A meta-analysis of the uncanny valley's independent and dependent variables. *ACM Transactions on Human-Robot Interaction*, 11.

[Dimberg et al., 2011] Dimberg, U., Andréasson, P., and Thunberg, M. (2011). Emotional empathy and facial reactions to facial expressions. *Journal of Psychophysiology*.

[Djebbara, 2022] Djebbara, Z. (2022). *Affordances in Everyday Life: A Multidisciplinary Collection of Essays*.

[Dlmberg, 1982] Dlmberg, U. (1982). Facial reactions to facial expressions.

[Doyle et al., 2019] Doyle, P. R., Edwards, J., Dumbleton, O., Clark, L., and Cowan, B. R. (2019). Mapping perceptions of humanness in intelligent personal assistant interaction. In *Proceedings of the 21st international conference on human-computer interaction with mobile devices and services*, pages 1–12.

[Draude, 2011] Draude, C. (2011). Intermediaries: Reflections on virtual humans, gender, and the uncanny valley. *AI and Society*, 26:319–327.

[Duan, 1996] Duan, C. (1996). The current state of empathy research. *Journal of Counseling Psychology*, 43:261–274.

[Duan, 2000] Duan, C. (2000). Being empathic:the role of motivation to empathize and the nature of target emotions. *Motivation and Emotion*, 24.

[Dukes et al., 2021] Dukes, D., Abrams, K., Adolphs, R., Ahmed, M. E., Beatty, A., Berridge, K. C., Broomhall, S., Brosch, T., Campos, J. J., Clay, Z., et al. (2021). The rise of affectivism. *Nature human behaviour*, 5(7):816–820.

[Durupinar and Kim, 2022] Durupinar, F. and Kim, J. (2022). Facial emotion recognition of virtual humans with different genders, races, and ages. Association for Computing Machinery, Inc.

[Döllinger et al., 2023] Döllinger, N., Wolf, E., Botsch, M., Latoschik, M. E., and Wienrich, C. (2023). Are embodied avatars harmful to our self-experience? the impact of virtual embodiment on body awareness. *Conference on Human Factors in Computing Systems - Proceedings*.

[Döllinger et al., 2022] Döllinger, N., Wolf, E., Mal, D., Erdmannsdörfer, N., Botsch, M., Latoschik, M. E., and Wienrich, C. (2022). Virtual reality for mind and body: Does the sense of embodiment towards a virtual body affect physical body awareness? pages 1–8. ACM.

[Eich et al., 2007] Eich, E., Ng, J. T. W., Macaulay, D., and Grebneva, A. D. P. (2007). *Combining Music With Thought to Change Mood*. Series in Affective Science. Oxford University Press, USA.

[EKMAN, 1999] EKMAN, P. (1999). Basic emotions. *Handbook of cognition and emotion*, pages 45–60.

[Ekman, 1999] Ekman, P. (1999). *Handbook of Cognition and Emotion*, chapter Basic Emotions, pages 45–60. John Wiley & Sons Ltd.

[Ekman and Friesen, 1978] Ekman, P. and Friesen, W. V. (1978). *Facial action coding system*. Consulting Psychologists Press, Stanford University.

[Ethofer et al., 2009] Ethofer, T., Van De Ville, D., Scherer, K., and Vuilleumier, P. (2009). Decoding of emotional information in voice-sensitive cortices. *Current biology*, 19(12):1028–1033.

[Fehr and Russell, 1984] Fehr, B. and Russell, J. A. (1984). Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General*, 113:464–486.

[Feldman et al., 2009] Feldman, N. H., Griffiths, T. L., and Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, 116:752–782.

[Ferrer et al., 2015] Ferrer, R. A., Grenen, E. G., and Taber, J. M. (2015). Effectiveness of internet-based affect induction procedures: A systematic review and meta-analysis. *Emotion*, 15:752–762.

[Ferstl and McDonnell, 2018] Ferstl, Y. and McDonnell, R. (2018). A perceptual study on the manipulation of facial features for trait portrayal in virtual agents. pages 281–288. Association for Computing Machinery (ACM).

[Ferstl et al., 2021a] Ferstl, Y., Thomas, S., Guiard, C., Ennis, C., and McDonnell, R. (2021a). Human or robot? investigating voice, appearance and gesture motion realism of conversational social agents. In *Proceedings of the 21th ACM International Conference on Intelligent Virtual Agents*, pages 76–83.

[Ferstl et al., 2021b] Ferstl, Y., Thomas, S., Guiard, C., Ennis, C., and McDonnell, R. (2021b). Human or robot?: Investigating voice, appearance and gesture motion realism of conversational social agents. pages 76–83. Association for Computing Machinery, Inc.

[Fink and Matts, 2008] Fink, B. and Matts, P. J. (2008). The effects of skin colour distribution and topography cues on the perception of female facial age and health. *Journal of the European Academy of Dermatology and Venereology*, 22(4):493–498.

[Fox et al., 2015] Fox, J., Ahn, S. J. G., Janssen, J. H., Yeykelis, L., Segovia, K. Y., and Bailenson, J. N. (2015). Avatars versus agents: A meta-analysis quantifying the effect of agency on social influence. *Human-Computer Interaction*, 30:401–432.

[Fox et al., 2013] Fox, J., Bailenson, J. N., and Tricase, L. (2013). The embodiment of sexualized virtual selves: The proteus effect and experiences of self-objectification via avatars. *Computers in Human Behavior*, 29:930–938.

[Freud, 2017] Freud, S. (2017). The uncanny. In *Romantic Writings*, pages 318–325. Routledge.

[Fribourg et al., 2018] Fribourg, R., Argelaguet, F., Hoyet, L., and Lécuyer, A. (2018). Studying the sense of embodiment in vr shared experiences. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 273–280.

[Fribourg et al., 2020] Fribourg, R., Argelaguet, F., Lécuyer, A., and Hoyet, L. (2020). Avatar and sense of embodiment: Studying the relative preference between appearance, control and point of view. *IEEE Transactions on Visualization and Computer Graphics*, 26:2062–2072.

[Fridlund AJ, 1986] Fridlund AJ, C. J. (1986). Guidelines for human electromyographic research. psychophysiology. *Psychophysiology.*, 23:567–89.

[Friston, 2022] Friston, K. (2022). *Affordance and Active Inference*, pages 211–219. Springer International Publishing, Cham.

[Gallagher, 2008] Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, 17:535–543.

[Gallagher, 2012] Gallagher, S. (2012). Empathy, simulation, and narrative. *Science in context*, 25(3):355–381.

[Gambino et al., 2020] Gambino, A., Fox, J., and Ratan, R. A. (2020). Building a stronger casa: Extending the computers are social actors paradigm. *Human-Machine Communication*, 1:71–85.

[Gao et al., 2017] Gao, X., Weng, L., Zhou, Y., and Yu, H. (2017). The influence of empathy and morality of violent video game characters on gamers' aggression. *Frontiers in Psychology*, 8.

[Garau et al., 2005] Garau, M., Slater, M., Pertaub, D. P., and Razzaque, S. (2005). The responses of people to virtual humans in an immersive virtual environment. volume 14, pages 104–116.

[Gendron and Barrett, 2018] Gendron, M. and Barrett, L. F. (2018). Emotion perception as conceptual synchrony. *Emotion Review*, 10(2):101–110.

[Gerrards-Hesse et al., 1994] Gerrards-Hesse, A., Spies, K., and Hesse, F. (1994). Experimental inductions of emotional states and their effectiveness: A review. *British Journal of Psychology*, 85:55–78.

[Gibson, 2000] Gibson, E. J. (2000). Where is the information for affordances? *Ecological Psychology*, 12:53–56.

[Gibson, 1966] Gibson, J. (1966). *The senses considered as perceptual systems*. Houghton Mifflin, Boston.

[Golan and Hill, 2006] Golan, Ofer, B.-C. S. and Hill, J. (2006). The cambridge mindreading (cam) face-voice battery: Testing complex emotion recognition in adults with and without asperger syndrome. *Journal of Autism and Developmental Disorders*, 36(2):169–183.

[Goldman, 2002] Goldman, A. I. (2002). Simulation theory and mental concepts. *Simulation and knowledge of action*, 119.

[Golland et al., 2018] Golland, Y., Hakim, A., Aloni, T., Schaefer, S., and Levit-Binnun, N. (2018). Affect dynamics of facial emg during continuous emotional experiences. *Biological Psychology*, 139:47–58.

[Gong and Nass, 2007] Gong, L. and Nass, C. (2007). When a talking-face computer agent is half-human and half-humanoid: Human identity and consistency preference. *Human Communication Research*, 33:163–193.

[Gonzalez-Franco and Lanier, 2017] Gonzalez-Franco, M. and Lanier, J. (2017). Model of illusions and virtual reality. *Frontiers in psychology*, 8:1125.

[Gonzalez-Franco and Peck, 2018] Gonzalez-Franco, M. and Peck, T. C. (2018). Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI*, 5:74.

[Gonzalez-Franco et al., 2020] Gonzalez-Franco, M., Steed, A., Hoogendyk, S., and Ofek, E. (2020). Using facial animation to increase the enfacement illusion and avatar self-identification. *IEEE Transactions on Visualization and Computer Graphics*, 26:2023–2029.

[Gorisse et al., 2019] Gorisse, G., Christmann, O., Houzangbe, S., and Richir, S. (2019). From robot to virtual doppelganger: Impact of visual fidelity of avatars controlled in third-person perspective on embodiment and behavior in immersive virtual environments. *Frontiers in Robotics and AI*, 6:8.

[Gosling et al., 2003] Gosling, S. D., Rentfrow, P. J., and Swann Jr, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in personality*, 37(6):504–528.

[Guadagno et al., 2007] Guadagno, R. E., Blascovich, J., Bailenson, J. N., and Mccall, C. (2007). Virtual humans and persuasion: The effects of agency and behavioral realism. *Media Psychology*, 10:1–22.

[Gulz and Haake, 2006] Gulz, A. and Haake, M. (2006). Visual design of virtual pedagogical agents: Naturalism versus stylization in static appearance. In *Proceedings of the 3rd International Design and Engagability Conference@ NordiChi 2006*.

[Guo et al., 2024] Guo, S., Choi, M., Kao, D., and Mousas, C. (2024). Collaborating with my doppelgänger: The effects of self-similar appearance and voice of a virtual character during a jigsaw puzzle co-solving task. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 7(1):1–23.

[Gupta et al., 2017] Gupta, N., Singh, A., and Butail, S. (2017). The effect of instructional priming on postural responses to virtual crowds. pages 1–8.

[Halan et al., 2015] Halan, S., Sia, I., Crary, M., and Lok, B. (2015). Exploring the effects of healthcare students creating virtual patients for empathy training. volume 9238, pages 239–249. Springer Verlag.

[Ham et al., 2024] Ham, J., Li, S., Looi, J., and Eastin, M. S. (2024). Virtual humans as social actors: Investigating user perceptions of virtual humans' emotional expression on social media. *Computers in Human Behavior*, 155:108161.

[Hanson et al., 2005] Hanson, D., Olney, A., Pereira, I. A., and Zielke, M. (2005). Upending the uncanny valley.

[Harmon-Jones et al., 2017] Harmon-Jones, E., Harmon-Jones, C., and Summerell, E. (2017). On the importance of both dimensional and discrete models of emotion. *Behavioral sciences*, 7(4):66.

[Hassouneh et al., 2020] Hassouneh, A., Mutawa, A., and Murugappan, M. (2020). Development of a real-time emotion recognition system using facial expressions and eeg based on machine learning and deep neural network methods. *Informatics in Medicine Unlocked*, 20:100372.

[Hawes and Arya, 2021] Hawes, D. and Arya, A. (2021). Vr-based student priming to reduce anxiety and increase cognitive bandwidth.

[Heide et al., 2013] Heide, B. V. D., Schumaker, E. M., Peterson, A. M., and Jones, E. B. (2013). The proteus effect in dyadic communication: Examining the effect of avatar appearance in computer-mediated dyadic interaction. *Communication Research*, 40(6):838–860.

[Heidicker et al., 2017] Heidicker, P., Langbehn, E., and Steinicke, F. (2017). Influence of avatar appearance on presence in social vr. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 233–234, Los Alamitos, CA, USA. IEEE Computer Society.

[Hershler and Hochstein, 2006] Hershler, O. and Hochstein, S. (2006). With a careful look: Still no low-level confound to face pop-out. *Vision Research*, 46:3028–3035.

[Hessels et al., 2016] Hessels, R. S., Kemner, C., van den Boomen, C., and Hooge, I. T. (2016). The area-of-interest problem in eyetracking research: A noise-robust solution for face and sparse stimuli. *Behavior Research Methods*, 48:1694–1712.

[Ho and MacDorman, 2010a] Ho, C. C. and MacDorman, K. F. (2010a). Revisiting the uncanny valley theory: Developing and validating an alternative to the godspeed indices. *Computers in Human Behavior*, 26:1508–1518.

[Ho and MacDorman, 2010b] Ho, C.-C. and MacDorman, K. F. (2010b). Revisiting the uncanny valley theory: Developing and validating an alternative to the godspeed indices. *Computers in Human Behavior*, 26(6):1508–1518.

[Ho and MacDorman, 2017] Ho, C. C. and MacDorman, K. F. (2017). Measuring the uncanny valley effect: Refinements to indices for perceived humanness, attractiveness, and eeriness. *International Journal of Social Robotics*, 9:129–139.

[Ho and Ng, 2020] Ho, J. C. and Ng, R. (2020). Perspective-taking of non-player characters in prosocial virtual reality games: Effects on closeness, empathy, and game immersion. *Behaviour and Information Technology*.

[Hoffman, 2024] Hoffman, D. D. (2024). Spacetime is doomed: Time is an artifact. *Timing Time Perception*, pages 1–3.

[Hoffman et al., 2015] Hoffman, D. D., Singh, M., and Prakash, C. (2015). The interface theory of perception. *Psychonomic Bulletin and Review*, 22:1480–1506.

[Holden et al., 2016] Holden, D., Saito, J., and Komura, T. (2016). A deep learning framework for character motion synthesis and editing. *ACM Transactions on Graphics (TOG)*, 35(4):1–11.

[Holland et al., 2021] Holland, A. C., O'Connell, G., and Dziobek, I. (2021). Facial mimicry, empathy, and emotion recognition: a meta-analysis of correlations. *Cognition and Emotion*, 35:150–168.

[Hu et al., 2019] Hu, P., Wang, K., and Liu, J. (2019). Speaking and listening: Mismatched human-like conversation qualities undermine social perception and trust in ai-based voice assistants. In *PACIS*.

[Hufendiek, 2017] Hufendiek, R. (2017). Affordances and the normativity of emotions. *Synthese*, 194:4455–4476.

[ichiro Seyama and Nagayama, 2007] ichiro Seyama and Nagayama, R. S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces.

[IJsselsteijn et al., 2006] IJsselsteijn, W. A., de Kort, Y. A. W., and Haans, A. (2006). Is this my hand i see before me? the rubber hand illusion in reality, virtual reality, and mixed reality. *Presence: Teleoperators and Virtual Environments*, 15(4):455–464.

[Ito et al., 1998] Ito, T. A., Larsen, J. T., Smith, N. K., and Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations. *Journal of personality and social psychology*, 75:887–900.

[Izard, 2010] Izard, C. E. (2010). The many meanings/aspects of emotion: Definitions, functions, activation, and regulation.

[Jardine and Szanto, 2017] Jardine, J. and Szanto, T. (2017). Empathy in the phenomenological tradition. In *The Routledge Handbook of Philosophy of Empathy*, pages 86–97. Routledge.

[Jensen and Pedersen, 2016] Jensen, T. W. and Pedersen, S. B. (2016). Affect and affordances–the role of action and emotion in social interaction. *Cognitive Semiotics*, 9(1):79–103.

[Jeunet et al., 2018] Jeunet, C., Albert, L., Sanz, F. A., Lécuyer, A., and Argelaguet, F. (2018). "do you feel in control?": Towards novel approaches to characterise, manipulate and measure the sense of agency in virtual environments. *IEEE Transactions on Visualization and Computer Graphics, Institute of Electrical and Electronics Engineers*, 24:1486–1495.

[John et al., 1999] John, O. P., Srivastava, S., et al. (1999). The big-five trait taxonomy: History, measurement, and theoretical perspectives.

[Jolien Van der Graaff and Branje, 2016] Jolien Van der Graaff, Wim Meeus, M. d. W. A. v. B. P. A. C. v. L. H. M. K. and Branje, S. (2016). Motor, affective and cognitive empathy in adolescence: Interrelations between facial electromyography and self-reported trait and state measures. *Cognition and Emotion*, 30(4):745–761.

[Jonell et al., 2020] Jonell, P., Kucherenko, T., Torre, I., and Beskow, J. (2020). Can we trust online crowdworkers?: Comparing online and offline participants in a preference test of virtual agents. Association for Computing Machinery, Inc.

[Joyal et al., 2014] Joyal, C. C., Jacob, L., Cigna, M. H., Guay, J. P., and Renaud, P. (2014). Virtual faces expressing emotions: An initial concomitant and construct validity study. *Frontiers in Human Neuroscience*, 8:1–6.

[Julie N. Buchan and Munhall, 2007] Julie N. Buchan, M. P. and Munhall, K. G. (2007). Spatial statistics of gaze fixations during dynamic face processing. *Social Neuroscience*, 2(1):1–13.

[Jung et al., 2018] Jung, S., Wisniewski, P. J., and Hughes, C. E. (2018). In limbo: The effect of gradual visual transition between real and virtual on virtual body ownership illusion and presence. pages 267–272.

[Jung et al., 2011] Jung, Y., Kuijper, A., Fellner, D., Kipp, M., Miksatko, J., Gratch, J., and Thalmann, D. (2011). Believable virtual characters in human-computer dialogs: State of the art report.

[Kahn et al., 2007] Kahn, J. H., Tobin, R. M., Massey, A. E., and Anderson, J. A. (2007). Measuring emotional expression with the linguistic inquiry and word count. *The American Journal of Psychology*, 120(2):263–286.

[Kaneko and Kameoka, 2018] Kaneko, T. and Kameoka, H. (2018). Cyclegan-vc: Non-parallel voice conversion using cycle-consistent adversarial networks. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 2100–2104. IEEE.

[Kang et al., 2008] Kang, S.-H., Gratch, J., Wang, N., and Watt, J. H. (2008). Agreeable people like agreeable virtual humans. In *Intelligent Virtual Agents: 8th International Conference, IVA 2008, Tokyo, Japan, September 1-3, 2008. Proceedings 8*, pages 253–261. Springer.

[Kao et al., 2021] Kao, D., Ratan, R., Mousas, C., and Magana, A. J. (2021). The effects of a self-similar avatar voice in educational games. volume 5. Association for Computing Machinery.

[Karnaze, 2013] Karnaze, M. M. (2013). A constructivist approach to defining human emotion: From george kelly to rue cromwell. *Journal of Constructivist Psychology*, 26:194–201.

[Katifori et al., 2022] Katifori, A., Lougiakis, C., and Roussou, M. (2022). Exploring the effect of personality traits in vr interaction: The emergent role of perspective-taking in task performance. *Frontiers in Virtual Reality*, 3:19.

[Kawahara et al., 2010] Kawahara, H., Morise, M., Takahashi, T., Banno, H., Nisimura, R., and Irino, T. (2010). Simplification and extension of non-periodic excitation source representations for high-quality speech manipulation systems. pages 38–41.

[Kawahara et al., 2008] Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., and Banno, H. (2008). Tandem-straight: A temporally stable power spectral

representation for periodic signals and applications to interference-free spectrum, f0, and aperiodicity estimation. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3933–3936.

[Kilteni et al., 2012] Kilteni, K., Groten, R., and Slater, M. (2012). The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21:373–387.

[Kim et al., 2020] Kim, C. S., Jung, M., Kim, S. Y., and Kim, K. (2020). Controlling the sense of embodiment for virtual avatar applications: Methods and empirical study. *JMIR Serious Games 2020;8(3):e21879 https://games.jmir.org/2020/3/e21879*, 8:e21879.

[Kiverstein, 2015] Kiverstein, J. (2015). Empathy and the responsiveness to social affordances. *Consciousness and Cognition*, 36:532–542.

[Kiverstein and Rietveld, 2012] Kiverstein, J. and Rietveld, E. (2012). Dealing with context through action-oriented predictive processing. *Frontiers in psychology*, 3:34012.

[Klatzky and Creswell, 2014] Klatzky, R. L. and Creswell, J. D. (2014). An intersensory interaction account of priming effects-and their absence. *Perspectives on Psychological Science*, 9:49–58.

[Kleinginna and Kleinginna, 1981] Kleinginna, P. R. and Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition.

[Kleinsmith et al., 2015] Kleinsmith, A., Rivera-Gutierrez, D., Finney, G., Cendan, J., and Lok, B. (2015). Understanding empathy training with virtual patients. *Computers in Human Behavior*, 52:151–158.

[Koda and Ishida, 2006] Koda, T. and Ishida, T. (2006). Cross-cultural study of avatar expression interpretations. *International Symposium on Applications and the Internet*.

[Koda and Oguri, 2019] Koda, T. and Oguri, R. (2019). Analysis of the effects of appearances of avatars on user's self-evaluation of extroversion. volume 1, pages 232–237. SciTePress.

[Kokkinara and McDonnell, 2015] Kokkinara, E. and McDonnell, R. (2015). Animation realism affects perceived character appeal of a self-virtual face. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games*, MIG '15, page 221–226.

[Kokkinara and Slater, 2014] Kokkinara, E. and Slater, M. (2014). Measuring the effects through time of the influence of visuomotor and visuotactile synchronous stimulation on a virtual body ownership illusion. *Perception*, 43:43–58.

[Kossaifi et al., 2019] Kossaifi, J., Walecki, R., Panagakis, Y., Shen, J., Schmitt, M., Ringeval, F., Han, J., Pandit, V., Toisoul, A., Schuller, B., Star, K., Hajiyev, E., and Pantic, M. (2019). Sewa db: A rich database for audio-visual emotion and sentiment research in the wild.

[Krum et al., 2018] Krum, D. M., Kang, S. H., and Phan, T. (2018). Influences on the elicitation of interpersonal space with virtual humans. *25th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2018 - Proceedings*, pages 223–229.

[Krystallidou et al., 2018] Krystallidou, D., Remael, A., de Boe, E., Hendrickx, K., Tsakitzidis, G., van de Geuchte, S., and Pype, P. (2018). Investigating empathy in interpreter-mediated simulated consultations: An explorative study. *Patient Education and Counseling*, 101:33–42.

[Kühne et al., 2020] Kühne, K., Fischer, M. H., and Zhou, Y. (2020). The human takes it all: Humanlike synthesized voices are perceived as less eerie and more likable. evidence from a subjective ratings study. *Frontiers in neurorobotics*, 14:593732.

[Kunz, 2015] Kunz, S. (2015). The problem of realism in animated characters: has the uncanny valley been crossed?

[Kwon et al., 2009] Kwon, J. H., Alan, C., Czanner, S., Czanner, G., and Powell, J. (2009). A study of visual perception: Social anxiety and virtual realism. In *Proceedings of the 25th spring conference on computer graphics*, pages 167–172.

[Kätsyri et al., 2019] Kätsyri, J., de Gelder, B., and Takala, T. (2019). Virtual faces evoke only a weak uncanny valley effect: An empirical investigation with controlled virtual face images. *Perception*, 48:968–991.

[Kätsyri et al., 2015] Kätsyri, J., Förger, K., Mäkäräinen, M., and Takala, T. (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.

[Kätsyri et al., 2017a] Kätsyri, J., Mäkäräinen, M., and Takala, T. (2017a). Testing the 'uncanny valley' hypothesis in semirealistic computer-animated film characters: An empirical evaluation of natural film stimuli. *International Journal of Human Computer Studies*, 97:149–161.

[Kätsyri et al., 2017b] Kätsyri, J., Mäkäräinen, M., and Takala, T. (2017b). Testing the 'uncanny valley' hypothesis in semirealistic computer-animated film characters: An empirical evaluation of natural film stimuli. *International Journal of Human Computer Studies*, 97:149–161.

[Künecke et al., 2014] Künecke, J., Hildebrandt, A., Recio, G., Sommer, W., and Wilhelm, O. (2014). Facial emg responses to emotional expressions are related to emotion perception ability. *PLoS ONE*, 9.

[Lange et al., 2022] Lange, J., Heerdink, M. W., and Van Kleef, G. A. (2022). Reading emotions, reading people: Emotion perception and inferences drawn from perceived emotions. *Current opinion in psychology*, 43:85–90.

[Larsen et al., 2003] Larsen, J. T., Norris, C. J., and Cacioppo, J. T. (2003). Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology*, 40(5):776–785.

[Lashley et al., 1951] Lashley, K. S. et al. (1951). The problem of serial order in behavior.

[Latoschik et al., 2017] Latoschik, M. E., Roth, D., Gall, D., Achenbach, J., Waltemate, T., and Botsch, M. (2017). The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM symposium on virtual reality software and technology*, pages 1–10.

[Lay et al., 2016] Lay, S., Brace, N., Pike, G., and Pollick, F. (2016). Circling around the uncanny valley: Design principles for research into the relation between human likeness and eeriness.

[LeDoux, 2012] LeDoux, J. (2012). Rethinking the emotional brain. *Neuron*, 73:653–676.

[Lee and Nass, 2003] Lee, K. M. and Nass, C. (2003). Designing social presence of social actors in human computer interaction.

[Lee et al., 2017] Lee, Y. I., Choi, Y., and Jeong, J. (2017). Character drawing style in cartoons on empathy induction: An eye-tracking and eeg study. *PeerJ*, 2017.

[Lindquist and Gendron, 2013] Lindquist, K. A. and Gendron, M. (2013). What's in a word? language constructs emotion perception. *Emotion Review*, 5(1):66–71.

[Lisa A. Elkin, 2021] Lisa A. Elkin, Matthew Kay, J. J. H. J. O. W. (2021). An aligned rank transform procedure for multifactor contrast tests. *arXiv*, page 11824.

[Llanes-Jurado et al., 2024] Llanes-Jurado, J., Gómez-Zaragozá, L., Minissi, M. E., Alcañiz, M., and Marín-Morales, J. (2024). Developing conversational virtual humans for social emotion elicitation based on large language models. *Expert Systems with Applications*, 246:123261.

[Longo and Bertenthal, 2009] Longo, M. R. and Bertenthal, B. I. (2009). Attention modulates the specificity of automatic imitation to human actors. *Experimental Brain Research*, 192:739–744.

[Loveys et al., 2020] Loveys, K., Sagar, M., and Broadbent, E. (2020). The effect of multimodal emotional expression on responses to a digital human during a self-disclosure conversation: a computational analysis of user language. *Journal of medical systems*, 44(9):143.

[Lundqvist, 1995] Lundqvist, L.-O. (1995). Facial e m g reactions to facial expressions: A cace of facial emotional contagion? *Scandinavian Journal of Psychology*, 36:130–141.

[Lush, 2020] Lush, P. (2020). Demand characteristics confound the rubber hand illusion. *Colabra:Psychology*, 6.

[Lush et al., 2020] Lush, P., Botan, V., Scott, R. B., Seth, A. K., Ward, J., and Dienes, Z. (2020). Trait phenomenological control predicts experience of mirror synaesthesia and the rubber hand illusion. *Nature Communications 2020 11:1*, 11:1–10.

[Ma and Hommel, 2013] Ma, K. and Hommel, B. (2013). The virtual-hand illusion: effects of impact and threat on perceived ownership and affective resonance. *Frontiers in Psychology*, 4:604.

[MacDorman and Chattopadhyay, 2016] MacDorman, K. F. and Chattopadhyay, D. (2016). Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition*, 146:190–205.

[MacDorman and Entezari, 2015] MacDorman, K. F. and Entezari, S. O. (2015). Individual differences predict sensitivity to the uncanny valley. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial SystemsInteraction Studies / Social Behaviour and Communication in Biological and Artificial SystemsInteraction Studies*, 16:141–172.

[Macdorman et al., 2008] Macdorman, K. F., Green, R. D., Ho, C.-C., and Koch, C. T. (2008). Too real for comfort? uncanny responses to computer generated faces.

[MacDorman et al., 2013] MacDorman, K. F., Srinivas, P., and Patel, H. (2013). The uncanny valley does not interfere with level 1 visual perspective taking. *Computers in Human Behavior*, 29(4):1671–1685.

[MaCdorman et al., 2013] MaCdorman, K. F., Srinivas, P., and Patel, H. (2013). The uncanny valley does not interfere with level 1 visual perspective taking. *Computers in Human Behavior*, 29:1671–1685.

[Maffiolo and Ochs, 2013] Maffiolo, V. and Ochs, M. (2013). The role of emotions in human-machine interaction. *Emotion-Oriented Systems*, pages 215–246.

[Mäkäräinen et al., 2015] Mäkäräinen, M., Kätsyri, J., Förger, K., and Takala, T. (2015). The funcanny valley: a study of positive emotional reactions to strangeness. In *Proceedings of the 19th international academic mindtrek conference*, pages 175–181.

[Makransky et al., 2019] Makransky, G., Wismer, P., and Mayer, R. E. (2019). A gender matching effect in learning with pedagogical agents in an immersive virtual reality science simulation. *Journal of Computer Assisted Learning*, 35(3):349–358.

[Marschner et al., 2015] Marschner, L., Pannasch, S., Schulz, J., and Graupner, S. T. (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*, 97:85–92.

[Martínez-Fiestas et al., 2015] Martínez-Fiestas, M., Del Jesus, M. I. V., Sánchez-Fernández, J., and Montoro-Rios, F. J. (2015). A psychophysiological approach for measuring response to messaging: How consumers emotionally process green advertising. *Journal of Advertising Research*, 55(2):192–205.

[Matravers, 2017] Matravers, D. (2017). Empathy in the aesthetic tradition. In Maibom, H., editor, *The Routledge Handbook of Philosophy of Empathy*, pages 77–85. Routledge, London and New York.

[Mattheyses and Verhelst, 2015] Mattheyses, W. and Verhelst, W. (2015). Audiovisual speech synthesis: An overview of the state-of-the-art. *Speech Communication*, 66:182–217.

[McCreery et al., 2012] McCreery, M. P., Krach, S. K., Schrader, P. G., and Boone, R. (2012). Defining the virtual self: Personality, behavior, and the psychology of embodiment. *Computers in Human Behavior*, 28:976–983.

[McDonnell, 2012] McDonnell, R. (2012). Appealing virtual humans. pages 102–111.

[McDonnell and Breidt, 2010] McDonnell, R. and Breidt, M. (2010). Face reality: Investigating the uncanny valley for virtual faces. In *ACM SIGGRAPH ASIA 2010 Sketches*, SA '10, New York, NY, USA. Association for Computing Machinery.

[McDonnell et al., 2012] McDonnell, R., Breidt, M., and Buelthoff, H. (2012). Render me real Investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics*, 31(4):91:1–91:11.

[McDonnell et al., 2009] McDonnell, R., Larkin, M., Hernández, B., Rudomin, I., and O'Sullivan, C. (2009). Eye-catching crowds: saliency based selective variation. *ACM Transactions on Graphics (TOG)*, 28(3):1–10.

[McDonnell and Mutlu, 2021] McDonnell, R. and Mutlu, B. (2021). Appearance. *The handbook on socially interactive agents: 20 years of research on embodied conversational agents, intelligent virtual agents, and social robotics volume 1: Methods, behavior, cognition*, pages 105–146.

[Mehling et al., 2009] Mehling, W. E., Gopisetty, V., Daubenmier, J., Price, C. J., and Hecht, F. M. (2009). Body awareness: Construct and self-report measures. *PLoS ONE*, 4:5614.

[Mehling et al., 2012] Mehling, W. E., Price, C., Daubenmier, J. J., Acree, M., Bartmess, E., and Stewart, A. (2012). The multidimensional assessment of interoceptive awareness (maia). *PloS one*, 7(11):e48230.

[Mennecke et al., 2010] Mennecke, B. E., Triplett, J. L., Hassall, L. M., and Conde, Z. J. (2010). Embodied social presence theory.

[Merleau-Ponty, 2004] Merleau-Ponty, M. (2004). *The world of perception*. Routledge.

[Meyer and Schvaneveldt, 1971] Meyer, D. E. and Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of experimental psychology*, 90 2:227–34.

[Minio-Paluello et al., 2020] Minio-Paluello, I., Porciello, G., Gandolfo, M., Boukarras, S., and Aglioti, S. M. (2020). The enfacement illusion boosts facial mimicry. *Cortex*, 123:113–123.

[Mitchell and Phillips, 2015] Mitchell, R. L. and Phillips, L. H. (2015). The overlapping relationship between emotion perception and theory of mind. *Neuropsychologia*, 70:1–10.

[Mitchell et al., 2011a] Mitchell, W. J., Szerszen, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., and MacDorman, K. F. (2011a). A mismatch in the human realism of face and voice produces an uncanny valley.

[Mitchell et al., 2011b] Mitchell, W. J., Szerszen Sr, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., and MacDorman, K. F. (2011b). A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception*, 2(1):10–12.

[Moody et al., 2007] Moody, E. J., McIntosh, D. N., Mann, L. J., and Weisser, K. R. (2007). More than mere mimicry? the influence of emotion on rapid facial reactions to faces. *Emotion*, 7(2):447.

[Moore, 2012] Moore, R. K. (2012). A bayesian explanation of the 'uncanny valley' effect and related psychological phenomena. *Scientific Reports*, 2.

[Moore, 2016] Moore, R. K. (2016). Is spoken language all-or-nothing? implications for future speech-based human-machine interaction.

[Mori, 1970] Mori, M. (1970). The uncanny valley. *Energy*, 7(4):33–35.

[Mori et al., 2012] Mori, M., MacDorman, K. F., and Kageki, N. (2012). The uncanny valley. *IEEE Robotics and Automation Magazine*, 19:98–100.

[Morie et al., 2005] Morie, J. F., Williams, J., Dozois, A., and Luigi, D.-P. (2005). The fidelity of feel: Emotional affordance in virtual environments. In *Proceedings of the 11th International Conference on Human-Computer Interaction, Las Vegas, NV*.

[Moser et al., 2007] Moser, E., Derntl, B., Robinson, S., Fink, B., Gur, R. C., and Grammer, K. (2007). Amygdala activation at 3t in response to human and avatar facial expressions of emotions. *Journal of Neuroscience Methods*, 161:126–133.

[Mothes-Lasch et al., 2011] Mothes-Lasch, M., Mentzel, H.-J., Miltner, W. H., and Straube, T. (2011). Visual attention modulates brain activation to angry voices. *Journal of Neuroscience*, 31(26):9594–9598.

[Mousas et al., 2018] Mousas, C., Anastasiou, D., and Spantidi, O. (2018). The effects of appearance and motion of virtual characters on emotional reactivity. *Computers in Human Behavior*, 86:99–108.

[Murata et al., 2016] Murata, A., Saito, H., Schug, J., Ogawa, K., and Kameda, T. (2016). Spontaneous facial mimicry is enhanced by the goal of inferring emotional states: Evidence for moderation of "automatic" mimicry by higher cognitive processes. *PloS one*, 11(4):e0153128.

[Mustafa et al., 2017] Mustafa, M., Guthe, S., Tauscher, J. P., Goesele, M., and Magnor, M. (2017). How human am i? eeg-based evaluation of animated virtual characters. volume 2017-May, pages 5098–5108. Association for Computing Machinery.

[Mäkäräinen et al., 2015] Mäkäräinen, M., Kätsyri, J., Förger, K., and Takala, T. (2015). The funcanny valley: A study of positive emotional reactions to strangeness. pages 175–181. Association for Computing Machinery, Inc.

[Mäkäräinen et al., 2014] Mäkäräinen, M., Kätsyri, J., and Takala, T. (2014). Exaggerating facial expressions: A way to intensify emotion or a way to the uncanny valley? *Cognitive Computation*, 6:708–721.

[Naranjo and Schmidt, 2012] Naranjo, J. R. and Schmidt, S. (2012). Is it me or not me? modulation of perceptual-motor awareness and visuomotor performance by mindfulness meditation. *BMC Neuroscience*, 13:1–17.

[Nass and Moon, 2000] Nass, C. and Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues*, 56(1):81–103.

[Nass et al., 1994] Nass, C., Steuer, J., and Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 72–78.

[Neta and Haas, 2019] Neta, M. and Haas, I. J. (2019). Movere: Characterizing the role of emotion and motivation in shaping human behavior. *Emotion in the Mind and Body*, pages 1–9.

[Newen et al., 2018] Newen, A., Gallagher, S., and Bruin, L. D. (2018). 4e cognition.

[Nowak and Biocca, 2003] Nowak, K. L. and Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence*, 12:481–494.

[Nowak and Fox, 2018] Nowak, K. L. and Fox, J. (2018). Avatars and computer-mediated communication: a review of the definitions, uses, and effects of digital representations. *Review of Communication Research*, 6:30–53.

[Nunez and Blake, 2003] Nunez, D. and Blake, E. (2003). Conceptual priming as a determinant of presence in virtual environments.

[Oh et al., 2018] Oh, C. S., Bailenson, J. N., and Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Front. Robot. AI 5: 114. doi: 10.3389/frobt*.

[Oh et al., 2016] Oh, S. Y., Bailenson, J., Krämer, N., and Li, B. (2016). Let the avatar brighten your smile: Effects of enhancing facial expressions in virtual environments. *PLoS ONE*, 11.

[Otono et al., 2023] Otono, R., Genay, A., Perusquía-Hernández, M., Isoyama, N., Uchiyama, H., Hachet, M., Lécuyer, A., and Kiyokawa, K. (2023). I'm transforming! effects of visual transitions to change of avatar on the sense of embodiment in ar. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pages 83–93.

[PALERMO and RHODES, 2007] PALERMO, R. and RHODES, G. (2007). Are you always on my mind? a review of how face perception and attention interact. *Neuropsychologia*, 45(1):75–92.

[Palmer and Clifford, 2020] Palmer, C. J. and Clifford, C. W. (2020). Face pareidolia recruits mechanisms for detecting human social attention. *Psychological Science*, 31:1001–1012.

[Panagiotopoulou et al., 2017] Panagiotopoulou, E., Filippetti, M. L., Tsakiris, M., and Fotopoulou, A. (2017). Affective touch enhances self-face recognition during multisensory integration. *Scientific Reports*, 7.

[Parmar et al., 2022] Parmar, D., Olafsson, S., Utami, D., Murali, P., and Bickmore, T. (2022). Designing empathic virtual agents: manipulating animation, voice, rendering, and empathy to create persuasive agents. *Autonomous Agents and Multi-Agent Systems*, 36:17.

[Parsons et al., 2017] Parsons, T. D., Gaggioli, A., and Riva, G. (2017). Virtual reality for research in social neuroscience. *Brain Sciences*, 7.

[Pelachaud, 2013] Pelachaud, C. (2013). *Emotion-oriented systems*. John Wiley & Sons.

[Pelachaud and Bilvi, 2003] Pelachaud, C. and Bilvi, M. (2003). Computational model of believable conversational agents. *Communication in multiagent systems*, 2650:300–317.

[Peña et al., 2009a] Peña, J., Hancock, J. T., and Merola, N. A. (2009a). The priming effects of avatars in virtual settings. *Communication Research*, 36:838–856.

[Peña et al., 2009b] Peña, J., Hancock, J. T., and Merola, N. A. (2009b). The priming effects of avatars in virtual settings. *Communication Research*, 36:838–856.

[Peña, 2011] Peña, J. F. (2011). Combiner l'influence de la perception et de l'opération d'avatars selon le modèle de l'automaticité des effets d'amorçage. *Communication Theory*, 21:150–168.

[Pino et al., 2021] Pino, M. C., Vagnetti, R., Valenti, M., and Mazza, M. (2021). Comparing virtual vs real faces expressing emotions in children with autism: An eye-tracking study. *Education and Information Technologies*, 26:5717–5732.

[Poggi and Pelachaud, 2001] Poggi, I. and Pelachaud, C. (2001). *Performative facial expressions in animated faces*, page 155–188. MIT Press, Cambridge, MA, USA.

[Polanyi, 1966] Polanyi, M. (1966). The logic of tacit inference. *Philosophy*, 41:1–18.

[Porciello et al., 2018] Porciello, G., Bufalari, I., Minio-Paluello, I., Di Pace, E., and Aglioti, S. M. (2018). The 'enfacement' illusion: A window on the plasticity of the self. *Cortex*, 104:261 – 275. Special Section: The body and cognition: the relation between body representations and higher level cognitive and social processes.

[Potard et al., 2016] Potard, B., Aylett, M., and Braude, D. (2016). Cross modal evaluation of high quality emotional speech synthesis with the virtual human toolkit. volume 10011, pages 190–197.

[Potdevin et al., 2020] Potdevin, D., Sabouret, N., and Clavel, C. (2020). Intimacy perception: Does the artificial or human nature of the interlocutor matter ? *International Journal of Human Computer Studies*, 142.

[Powell and Roberts, 2017] Powell, P. A. and Roberts, J. (2017). Situational determinants of cognitive, affective, and compassionate empathy in naturalistic digital interactions. *Computers in Human Behavior*, 68:137–148.

[Preston, 2007] Preston, S. D. (2007). A perception-action model for empathy. *Empathy in Mental Illness*, pages 428–447.

[Preston and de Waal, 2002] Preston, S. D. and de Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, 25:1–20.

[Prince et al., 2015] Prince, E. B., Martin, K. B., Messinger, D. S., and Allen, M. (2015). Facial action coding system. *Environmental Psychology & Nonverbal Behavior*, 1.

[Prochazkova and Kret, 2017] Prochazkova, E. and Kret, M. E. (2017). Connecting minds and sharing emotions through mimicry: A neurocognitive model of emotional contagion. *Neuroscience and Biobehavioral Reviews*, 80:99–114.

[Provoost et al., 2017] Provoost, S., Lau, H. M., Ruwaard, J., and Riper, H. (2017). Embodied conversational agents in clinical psychology: a scoping review. *Journal of medical Internet research*, 19(5):e151.

[Psotka and Davison, 1993] Psotka, J. and Davison, S. (1993). Cognitive factors associated with immersion in virtual environments. In *NASA. Johnson Space Center, Proceedings of the 1993 Conference on Intelligent Computer-Aided Training and Virtual Environment Technology*.

[Pulos et al., 2004] Pulos, S., Elison, J., and Lennon, R. (2004). The hierarchical structure of the interpersonal reactivity index. *Social Behavior & Personality: an international journal*, 32(4).

[Pütten et al., 2014] Pütten, A. M. R.-V. D., Schulte, F. P., Eimler, S. C., Sobieraj, S., Hoffmann, L., Maderwald, S., Brand, M., and Krämer, N. C. (2014). Investigations on empathy towards humans and robots using fmri. *Computers in Human Behavior*, 33:201–212.

[Qi et al., 2021] Qi, Y., Bruch, D., Krop, P., Herrmann, M. J., Latoschik, M. E., Deckert, J., and Hein, G. (2021). Social buffering of human fear is shaped by gender, social concern, and the presence of real vs virtual agents. *Translational psychiatry*, 11(1):641.

[Qu et al., 2014] Qu, C., Brinkman, W.-P., Ling, Y., Wiggers, P., and Heynderickx, I. (2014). Conversations with a virtual human: Synthetic emotions and human responses. *Computers in Human Behavior*, 34:58–68.

[Raamkumar and Yang, 2023] Raamkumar, A. S. and Yang, Y. (2023). Empathetic conversational systems: A review of current advances, gaps, and opportunities. *IEEE Transactions on Affective Computing*, 14(4):2722–2739.

[Raghavendra et al., 2010] Raghavendra, E. V., Vijayaditya, P., and Prahallad, K. (2010). Speech synthesis using artificial neural networks. In *2010 National Conference On Communications (NCC)*, pages 1–5. IEEE.

[Raiturkar et al.,] Raiturkar, P., Farid, H., and Jain, E. Identifying computer-generated portraits: an eye tracking study.

[Ratan, 2012] Ratan, R. (2012). Self-presence, explicated: Body, emotion, and identity extension into the virtual self.

[Ratan et al., 2020] Ratan, R., Beyea, D., Li, B. J., and Graciano, L. (2020). Avatar characteristics induce users' behavioral conformity with small-to-medium effect sizes: a meta-analysis of the proteus effect. *Media Psychology*, 23(5):651–675.

[Ravaja et al., 2018] Ravaja, N., Bente, G., Katsyri, J., Salminen, M., and Takala, T. (2018). Virtual character facial expressions influence human brain and facial emg

activity in a decision-making game. *IEEE Transactions on Affective Computing*, 9:285–298.

[Reader and Savin-Baden, 2020] Reader, J. and Savin-Baden, M. (2020). Ethical conundrums and virtual humans. *Postdigital Science and Education*, 2:289–301.

[Reaz et al., 2006] Reaz, M. B., Hussain, M. S., and Mohd-Yasin, F. (2006). Techniques of emg signal analysis: Detection, processing, classification and applications. *Biological Procedures Online*, 8:11–35.

[Reinhard et al., 2020] Reinhard, R., Shah, K. G., Faust-Christmann, C. A., and Lachmann, T. (2020). Acting your avatar's age: effects of virtual reality avatar embodiment on real life walking speed. *Media Psychology*, 23:293–315.

[Reisenzein, 2000] Reisenzein, R. (2000). Wundt's three-dimensional theory of emotion. In Balzer, W., Sneed, J. D., and Moulines, C. U., editors, *Structuralist Knowledge Representation: Paradigmatic Examples (Poznan Studies in the Philosophy of the Sciences and the Humanities, 75, 219-250)*, pages 75–219. Rodopi.

[Rekik et al., 2024] Rekik, R., Wuhrer, S., Hoyet, L., Zibrek, K., and Olivier, A.-H. (2024). Number 2 star-state of the art report. *EUROGRAPHICS 2024 A. Aristidou and R. McDonnell (Guest Editors) COMPUTER GRAPHICS forum*, 43.

[Ripley, 2006] Ripley, E. B. D. (2006). A review of paying research participants: It's time to move beyond the ethical debate. *Journal of Empirical Research on Human Research Ethics*, 1(4):9–19. PMID: 19385834.

[Riva, 2018] Riva, G. (2018). The neuroscience of body memory: From the self through the space to the others. *Cortex*, 104:241–260.

[Rodero and Lucas, 2023] Rodero, E. and Lucas, I. (2023). Synthetic versus human voices in audiobooks: The human emotional intimacy effect. *New Media & Society*, 25(7):1746–1764.

[Rodrigues et al., 2015] Rodrigues, S. H., Mascarenhas, S., Dias, J., and Paiva, A. (2015). A process model of empathy for virtual agents. *Interacting with Computers*, 27(4):371–391.

[Romano et al., 2014] Romano, D., Pfeiffer, C., Maravita, A., and Blanke, O. (2014). Illusory self-identification with an avatar reduces arousal responses to painful stimuli. *Behavioural Brain Research*, 261:275–281.

[Roseman, 1996] Roseman, I. J. (1996). Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory. *Cognition and Emotion*, 10(3):241–278.

[Roth et al., 2019] Roth, D., Bloch, C., Schmitt, J., Frischlich, L., Latoschik, M. E., and Bente, G. (2019). Perceived authenticity, empathy, and pro-social intentions evoked through avatar-mediated self-disclosures. pages 21–30. Association for Computing Machinery.

[Roth et al., 2018] Roth, D., Klelnbeck, C., and Feigl, T. (2018). Beyond replication: Augmenting social behaviors in multi-user virtual realities. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 215–222.

[Roth and Latoschik, 2020] Roth, D. and Latoschik, M. E. (2020). Construction of the virtual embodiment questionnaire (veq). *IEEE Transactions on Visualization and Computer Graphics*, 26:3546–3556.

[Ruimi et al., 2022] Ruimi, L., Hadash, Y., Tanay, G., and Bernstein, A. (2022). *State Mindfulness Scale (SMS)*, pages 1–16. Springer International Publishing, Cham.

[Rukavina S, 2016] Rukavina S, Gruss S, H. H. T. J. W. S. T. H. (2016). Affective computing and the impact of gender and age. *PLoS ONE*, 3.

[Rymarczyk et al., 2016] Rymarczyk, K., Łukasz Zurawski, Jankowiak-Siuda, K., and Szatkowska, I. (2016). Do dynamic compared to static facial expressions of happiness and anger reveal enhanced facial mimicry? *PLoS ONE*, 11.

[Sacau et al., 2008] Sacau, A., Laarni, J., and Hartmann, T. (2008). Influence of individual factors on presence. *Computers in Human Behavior*, 24(5):2255–2273.

[Sallnäs, 2010] Sallnäs, E.-L. (2010). Haptic feedback increases perceived social presence. In *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*, pages 178–185. Springer.

[Sander, 2013] Sander, D. (2013). Models of emotion. *The Cambridge handbook of human affective neuroscience*, pages 5–56.

[Saneyoshi et al., 2022] Saneyoshi, A., Okubo, M., Suzuki, H., Oyama, T., and Laeng, B. (2022). The other-race effect in the uncanny valley. *International Journal of Human-Computer Studies*, 166:102871.

[Sas and O'Hare, 2003] Sas, C. and O'Hare, G. M. (2003). Presence equation: An investigation into cognitive factors underlying presence.

[Sato et al., 2008] Sato, W., Fujimura, T., and Suzuki, N. (2008). Enhanced facial emg activity in response to dynamic facial expressions. *International Journal of Psychophysiology*, 70:70–74.

[Schilbach et al., 2005] Schilbach, L., Helmert, J. R., Mojzisch, A., Pannasch, S., Velichkovsky, B. M., and Vogeley, K. (2005). Neural correlates, visual attention and facial expression during social interaction with virtual others. In *Proceedings of the 27th Annual Conference of Cognitive Science Society, Stresa, Italy*, pages 21–23.

[Schindler et al., 2017] Schindler, S., Zell, E., Botsch, M., and Kissler, J. (2017). Differential effects of face-realism and emotion on event-related brain potentials and their implications for the uncanny valley theory. *Scientific Reports*, 7.

[Schirmer and Adolphs, 2017] Schirmer, A. and Adolphs, R. (2017). Emotion perception from face, voice, and touch: comparisons and convergence. *Trends in cognitive sciences*, 21(3):216–228.

[Schurgin et al., 2014] Schurgin, M., Nelson, J., Iida, S., Ohira, H., Chiao, J., and Franconeri, S. (2014). Eye movements during emotion recognition in faces. *Journal of vision*, 14(13):14–14.

[Scott et al., 2020] Scott, K. M., Ashby, S., and Hanna, J. (2020). "human, all too human": Noaa weather radio and the emotional impact of synthetic voices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–9, New York, NY, USA. Association for Computing Machinery.

[Seymour, 2019] Seymour, M. (2019). Facs at 40: facial action coding system panel. In *ACM SIGGRAPH 2019 Panels*, pages 1–2.

[Seymour et al., 2019] Seymour, M., Yuan, L., Dennis, A. R., and Riemer, K. (2019). *Crossing the Uncanny Valley? Understanding Affinity, Trustworthiness, and Preference for More Realistic Virtual Humans in Immersive Environments*.

[Sforza et al., 2010] Sforza, A., Bufalari, I., Haggard, P., and Aglioti, S. M. (2010). My face in yours: Visuo-tactile facial stimulation influences sense of identity. *Social Neuroscience*, 5:148–162.

[Shen, 2010] Shen, L. (2010). Mitigating psychological reactance: The role of message-induced empathy in persuasion. *Human Communication Research*, 36:397–422.

[Shenli Peng and Hu, 2021] Shenli Peng, L. Z. and Hu, P. (2021). Relating self–other overlap to ingroup bias in emotional mimicry. *Social Neuroscience*, 16(4):439–447. PMID: 34098832.

[Shi et al., 2023] Shi, Z., Chen, H., Velentza, A.-M., Liu, S., Dennler, N., O'Connell, A., and Mataric, M. (2023). Evaluating and personalizing user-perceived quality of text-to-speech voices for delivering mindfulness meditation with different physical embodiments. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pages 516–524.

[Sierra Rativa et al., 2020] Sierra Rativa, A., Postma, M., and Van Zaanen, M. (2020). The influence of game character appearance on empathy and immersion: Virtual non-robotic versus robotic animals. *Simulation & Gaming*, 51(5):685–711.

[Silvanto et al., 2015] Silvanto, J., Ainley, V., Farb, N., Mehling, W. E., Daubenmier, J., Price, C. J., Gard, T., Kerr, C., Dunn, B. D., Klein, A. C., and Paulus, M. P. (2015). Interoception, contemplative practice, and health. *Front. Psychol*, 6:763.

[Simon Kimmel and Boll., 2023] Simon Kimmel, Frederike Jung, A. M. W. H. and Boll., S. (2023). Let's face it: Infuence of facial expressions on social presence in collaborative virtual reality. *In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*, page 16.

[Skalski and Whitbred, 2010] Skalski, P. and Whitbred, R. (2010). Image versus sound: A comparison of formal feature effects on presence and video game enjoyment. *PsychNology Journal*, 8(1).

[Skarbez et al., 2017] Skarbez, R., Neyret, S., Brooks, F. P., Slater, M., and Whitton, M. C. (2017). A psychophysical experiment regarding components of the plausibility illusion. *IEEE transactions on visualization and computer graphics*, 23(4):1369–1378.

[Slaby, 2012] Slaby, J. (2012). Affective self-construal and the sense of ability. *Emotion Review*, 4(2):151–156.

[Slater, 2009a] Slater, M. (2009a). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364:3549–3557.

[Slater, 2009b] Slater, M. (2009b). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1535):3549–3557.

[Slater et al., 2006a] Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., Pistrang, N., and Sanchez-Vives, M. V. (2006a). A virtual reprise of the stanley milgram obedience experiments. *PloS one*, 1(1):e39.

[Slater et al., 2006b] Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., Pistrang, N., and Sanchez-Vives, M. V. (2006b). A virtual reprise of the stanley milgram obedience experiments. *PLOS ONE*, 1(1):1–10.

[Slater et al., 2018] Slater, M., Herbelin, B., McMahan, R. P., Bertrand, P., Zenasni, F., Guegan, J., Robieux, L., and McCall, C. A. (2018). Learning empathy through virtual reality: Multiple strategies for training empathy-related abilities using body ownership illusions in embodied virtual reality. *Article*, 5:1.

[Slater et al., 2009] Slater, M., Khanna, P., Mortensen, J., and Yu, I. (2009). Visual realism enhances realistic response in an immersive virtual environment. *IEEE computer graphics and applications*, 29(3):76–84.

[Slater and Burden, 2009] Slater, S. and Burden, D. (2009). Emotionally responsive robotic avatars as characters in virtual worlds. In *2009 Conference in Games and Virtual Worlds for Serious Applications*, pages 12–19. IEEE.

[Smolentsev et al., 2017] Smolentsev, A., Cornick, J. E., and Blascovich, J. (2017). Using a preamble to increase presence in digital virtual environments. *Virtual Real.*, 21(3):153–164.

[Solomon, 1999] Solomon, R. C. (1999). The philosophy of emotions. In Lewis, M. and Havil, J., editors, *Handbook of Emotions*. Guilford Press.

[Sonnby–Borgström, 2002] Sonnby–Borgström, M. (2002). Automatic mimicry reactions as related to differences in emotional empathy. *Scandinavian Journal of Psychology*, 43(5):433–443.

[Steckenfinger and Ghazanfar, 2009] Steckenfinger, S. A. and Ghazanfar, A. A. (2009). Monkey visual behavior falls into the uncanny valley.

[Steed et al., 2016] Steed, A., Pan, Y., Zisch, F., and Steptoe, W. (2016). The impact of a self-avatar on cognitive load in immersive virtual reality. pages 67–76.

[Stevenage and Neil, 2014] Stevenage, S. V. and Neil, G. J. (2014). Faces and seeing voices: The integration and interaction of face and voice processing. volume 54, pages 266–281. Ubiquity Press Ltd.

[Stuart, 2022] Stuart, S. (2022). *Eye tracking: Background, methods, and applications*, volume 183. Springer Nature.

[Suh et al., 2011] Suh, K.-S., Kim, H., and Suh, E. K. (2011). What if your avatar looks like you? dual-congruity perspectives for avatar use. *Source: MIS Quarterly*, 35:711–729.

[Tajadura-Jiménez et al., 2012] Tajadura-Jiménez, A., Longo, M. R., Coleman, R., and Tsakiris, M. (2012). The person in the mirror: Using the enfacement illusion to investigate the experiential structure of self-identification. *Consciousness and Cognition*, 21(4):1725–1738.

[Tanay and Bernstein, 2013] Tanay, G. and Bernstein, A. (2013). State mindfulness scale (sms): development and initial validation. *Psychological assessment*, 25(4):1286.

[Tarnowski et al., 2020] Tarnowski, P., Kołodziej, M., Majkowski, A., and Rak, R. J. (2020). Eye-tracking analysis for emotion recognition. *Computational intelligence and neuroscience*, 2020(1):2909267.

[Taylor, 2009] Taylor, P. (2009). *Text-to-speech synthesis*. Cambridge university press.

[Teper and Inzlicht, 2013] Teper, R. and Inzlicht, M. (2013). Meditation, mindfulness and executive control: The importance of emotional acceptance and brain-based performance monitoring. *Social Cognitive and Affective Neuroscience*, 8:85–92.

[Thézé et al., 2020] Thézé, R., Gadiri, M. A., Albert, L., Provost, A., Giraud, A. L., and Mégevand, P. (2020). Animated virtual characters to explore audio-visual speech in controlled and naturalistic environments. *Scientific Reports*, 10.

[Tinwell et al., 2011] Tinwell, A., Grimshaw, M., Nabi, D. A., and Williams, A. (2011). Facial expression of emotion and perception of the uncanny valley in virtual characters. *Computers in Human Behavior*, 27:741–749.

[Tolba et al., 2018] Tolba, R. M., Al-Arif, T., and Horbaty, E.-S. M. E. (2018). Realistic facial animation review: Based on facial action coding system.

[Tondu and Bardou, 2011] Tondu, B. and Bardou, N. (2011). A new interpretation of mori's uncanny valley for future humanoid robots. *International Journal of Robotics and Automation*, 26:337–348.

[Torre et al., 2018] Torre, I., Carrigan, E., McCabe, K., McDonnell, R., and Harte, N. (2018). Survival at the museum: A cooperation experiment with emotionally expressive virtual characters. In *Proceedings of the 2018 on International Conference on Multimodal Interaction*, pages 423–427. ACM.

[Torre et al., 2024] Torre, I., Holk, S., Yadollahi, E., Leite, I., McDonnell, R., and Harte, N. (2024). Smiling in the face and voice of avatars and robots: Evidence for a 'smiling mcgurk effect'. *IEEE Transactions on Affective Computing*, 15(2):393–404.

[Treves et al.,] Treves, I. N., Tello, L. Y., Davidson, R. J., and Goldberg, S. B. the relationship between mindfulness and objective measures of body awareness: A meta-analysis.

[Tsakiris et al., 2011] Tsakiris, M., Tajadura-Jiménez, A., and Costantini, M. (2011). Just a heartbeat away from one's body: interoceptive sensitivity predicts malleability of body-representations. *Proceedings of the Royal Society B: Biological Sciences*, 278:2470.

[Turk and Schroder, 2010] Turk, O. and Schroder, M. (2010). Evaluation of expressive speech synthesis with voice conversion and copy resynthesis techniques. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(5):965–973.

[Vallverdú and Trovato, 2016] Vallverdú, J. and Trovato, G. (2016). Emotional affordances for human–robot interaction. *Adaptive Behavior*, 24(5):320–334.

[Van Der Heide et al., 2013] Van Der Heide, B., Schumaker, E. M., Peterson, A. M., and Jones, E. B. (2013). The proteus effect in dyadic communication: Examining the effect of avatar appearance in computer-mediated dyadic interaction. *Communication Research*, 40(6):838–860.

[van Dijke et al., 2020] van Dijke, J., van Nistelrooij, I., Bos, P., and Duyndam, J. (2020). Towards a relational conceptualization of empathy. *Nursing Philosophy*, 21.

[van Loon et al., 2018a] van Loon, A., Bailenson, J., Zaki, J., Bostick, J., and Willer, R. (2018a). Virtual reality perspective-taking increases cognitive empathy for specific others. *PLoS ONE*, 13.

[van Loon et al., 2018b] van Loon, A., Bailenson, J., Zaki, J., Bostick, J., and Willer, R. (2018b). Virtual reality perspective-taking increases cognitive empathy for specific others. *PLoS ONE*, 13.

[Van Vugt et al., 2006] Van Vugt, H. C., Hoorn, J. F., Konijn, E. A., and de Bie Dimitriadou, A. (2006). Affective affordances: Improving interface character engagement through interaction. *International Journal of Human-Computer Studies*, 64(9):874–888.

[Varela, 1997] Varela, F. J. (1997). Patterns of life: Intertwining identity and cognition.

[Vehlen et al., 2021] Vehlen, A., Spenthof, I., Tönsing, D., Heinrichs, M., and Domes, G. (2021). Evaluation of an eye tracking setup for studying visual attention in face-to-face conversations. *Scientific Reports*, 11.

[Vo et al., 2012] Vo, M. L., Smith, T. J., Mital, P. K., and Henderson, J. M. (2012). Do the eyes really have it? dynamic allocation of attention when viewing moving faces. *Journal of Vision*, 12.

[Volante et al., 2016] Volante, M., Babu, S. V., Chaturvedi, H., Newsome, N., Ebrahimi, E., Roy, T., Daily, S. B., and Fasolino, T. (2016). Effects of virtual human appearance fidelity on emotion contagion in affective inter-personal simulations. *IEEE Transaction on Visualization and Computer Graphics*, 22(4):1326–1335.

[Volonte et al., 2019] Volonte, M., Anaraky, R. G., Knijnenburg, B. P., Duchowski, A. T., and Babu, S. V. (2019). Empirical evaluation of the interplay of emotion and visual attention in human-virtual human interaction. Association for Computing Machinery, Inc.

[Von Der Pütten et al., 2010] Von Der Pütten, A. M., Krämer, N. C., and Gratch, J. (2010). How our personality shapes our interactions with virtual characters-implications for research and development. In *Intelligent Virtual Agents: 10th International Conference, IVA 2010, Philadelphia, PA, USA, September 20-22, 2010. Proceedings 10*, pages 208–221. Springer.

[Wages et al.,] Wages, R., Grünvogel, S. M., and Grützmacher, B. How realistic is realism? considerations on the aesthetics of computer games. *Proceedings. Lecture Notes in Computer Science*, 3166:216–225.

[Wallach et al., 2010] Wallach, H. S., Safir, M. P., and Samana, R. (2010). Personality variables and presence. *Virtual Reality*, 14:3–13.

[Waltemate et al., 2018] Waltemate, T., Gall, D., Roth, D., Botsch, M., and Latoschik, M. E. (2018). The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Transactions on Visualization and Computer Graphics*, 24:1643–1652.

[Wang and Doube, 2011] Wang, N. and Doube, W. (2011). How real is reality? a perceptually motivated system for quantifying visual realism in digital images. volume 2, pages 141–149.

[Wang et al., 2015] Wang, S., Lilienfeld, S. O., and Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology*, 19:393–407.

[Wang et al., 2019] Wang, T., Sato, Y., Otsuki, M., Kuzuoka, H., and Suzuki, Y. (2019). Effect of full body avatar in augmented reality remote collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1221–1222, Los Alamitos, CA, USA. IEEE Computer Society.

[Watson et al., 1988] Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of personality and social psychology*, 54(6):1063.

[Weibel et al.,] Weibel, D., Wissmath, B. U., and Mast, F. W. Immersion in mediated environments: The role of personality traits.

[Weibel et al., 2010] Weibel, D., Wissmath, B. U., and Mast, F. W. (2010). Immersion in mediated environments: The role of personality traits. *CYBERPSYCHOLOGY, BEHAVIOR, AND SOCIAL NETWORKING*, 13.

[Weingarten et al., 2015] Weingarten, E., Chen, Q., McAdams, M., Yi, J., Hepler, J., and Albarracín, D. (2015). From primed concepts to action: A meta-analysis of the behavioral effects of incidentally presented words. *Psychological Bulletin*, 142:472.

[Weise et al., 2011] Weise, T., Bouaziz, S., Li, H., and Pauly, M. (2011). Realtime performance-based facial animation. *ACM transactions on graphics (TOG)*, 30(4):1–10.

[Wienrich et al., 2018] Wienrich, C., Schindler, K., Döllinqer, N., Kock, S., and Traupe, O. (2018). Social presence and cooperation in large-scale multi-user virtual reality - the relevance of social interdependence for location-based environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 207–214.

[Wilkie and Diddams, 2021] Wilkie, I. and Diddams, N. (2021). Waves of laughter: comic surfing on bergson's mechanical inelasticity. *Comedy Studies*, 12(1):91–103.

[Wilson, 2002] Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4):625—636.

[Wiltshire et al., 2015] Wiltshire, T. J., Lobato, E. J., McConnell, D. S., and Fiore, S. M. (2015). Prospects for direct social perception: A multi-theoretical integration to further the science of social cognition. *Frontiers in Human Neuroscience*, 8.

[Wisessing et al., 2020] Wisessing, P., Zibrek, K., Cunningham, D. W., Dingliana, J., and McDonnell, R. (2020). Enlighten me: Importance of brightness and shadow for character emotion and appeal. *ACM Transactions on Graphics*, 39(3).

[Yamada et al., 2013] Yamada, Y., Kawabe, T., and Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the "uncanny valley" phenomenon. *Japanese psychological research*, 55(1):20–32.

[Yee and Bailenson, 2007] Yee, N. and Bailenson, J. (2007). The proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research*, 33:271–290.

[Yee and Bailenson, 2009] Yee, N. and Bailenson, J. N. (2009). The difference between being and seeing: The relative contribution of self-perception and priming to behavioral changes via digital self-representation. *Media Psychology*, 12(2):195–209.

[Yoon et al., 2019] Yoon, B., Kim, H., Lee, G. A., Billinghurst, M., and Woo, W. (2019). The effect of avatar appearance on social presence in an augmented reality remote collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 547–556.

[Yoshimura et al., 1999] Yoshimura, T., Tokuda, K., Masuko, T., Kobayashi, T., and Kitamura, T. (1999). Simultaneous modeling of spectrum, pitch and duration in hmm-based speech synthesis. In *Sixth European conference on speech communication and technology*.

[Zadra and Clore, 2011] Zadra, J. R. and Clore, G. L. (2011). Emotion and perception: The role of affective information. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2:676–685.

[Zahavi, 2011] Zahavi, D. (2011). Empathy and direct social perception: A phenomenological proposal. *Review of Philosophy and Psychology*, 2:541–558.

[Zalake et al., 2024] Zalake, M., Gomes De Siqueira, A., Vaddiparti, K., Antonenko, P., and Lok, B. (2024). Evaluating the effect of adapting virtual humans based on individual differences in users. In Duffy, V. G., editor, *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management*, pages 405–423, Cham. Springer Nature Switzerland.

[Zebrowitz et al., 2011] Zebrowitz, L. A., Bronstad, P. M., and Montepare, J. M. (2011). An ecological theory of face perception. *The Science of Social Vision: The Science of Social Vision*, 7(1).

[Zebrowitz and Montepare, 2008] Zebrowitz, L. A. and Montepare, J. M. (2008). Social psychological face perception: Why appearance matters. *Social and personality psychology compass*, 2(3):1497–1517.

[Zell et al., 2015a] Zell, E., Aliaga, C., Jarabo, A., Zibrek, K., Gutierrez, D., McDonnell, R., and Botsch, M. (2015a). To stylize or not to stylize? the effect of shape and material stylization on the perception of computer-generated faces. *ACM Transactions on Graphics (TOG)*, 34(6):1–12.

[Zell et al., 2015b] Zell, E., Aliaga, C., Jarabo, A., Zibrek, K., Gutierrez, D., McDonnell, R., and Botsch, M. (2015b). To stylize or not to stylize?: The effect of shape and material stylization on the perception of computer-generated faces. *ACM Transactions on Graphics*, 34(6):184:1–184:12.

[Zen et al., 2013] Zen, H., Senior, A., and Schuster, M. (2013). Statistical parametric speech synthesis using deep neural networks. In *2013 ieee international conference on acoustics, speech and signal processing*, pages 7962–7966. IEEE.

[Zhou and Meng, 2020] Zhou, L. F. and Meng, M. (2020). Do you see the "face"? individual differences in face pareidolia. *Journal of Pacific Rim Psychology*, 14.

[Zhou et al., 2019] Zhou, M. X., Mark, G., Li, J., and Yang, H. (2019). Trusting virtual agents: The effect of personality. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 9(2-3):1–36.

[Zibrek et al., 2021] Zibrek, K., Cabral, J., and McDonnell, R. (2021). Does synthetic voice alter social response to a photorealistic character in virtual reality? In *Proceedings of Motion Interaction and Games (to appear)*.

[Zibrek et al., 2017] Zibrek, K., Kokkinara, E., and McDonnell, R. (2017). Don't stand so close to me: investigating the effect of control on the appeal of virtual humans using immersion and a proximity-based behavioral task. In *Proceedings of the ACM Symposium on Applied Perception*, page 3. ACM.

[Zibrek et al., 2018] Zibrek, K., Kokkinara, E., and McDonnell, R. (2018). The effect of realistic appearance of virtual characters in immersive environments - does the character's personality play a role? *IEEE Transactions on Visualization and Computer Graphics*, 24:1681–1690.

[Zibrek et al., 2019] Zibrek, K., Martin, S., and McDonnell, R. (2019). Is photorealism important for perception of expressive virtual humans in virtual reality? *ACM Trans. Appl. Percept.*, 16(3).

[Zibrek and McDonnell, 2014] Zibrek, K. and McDonnell, R. (2014). Does render style affect perception of personality in virtual humans? pages 111–115. Association for Computing Machinery.

[Zibrek and McDonnell, 2019] Zibrek, K. and McDonnell, R. (2019). Social presence and place illusion are affected by photorealism in embodied vr. In *Motion, interaction and games*, pages 1–7.

[Zibrek et al., 2020] Zibrek, K., Niay, B., Olivier, A.-H., Hoyet, L., Pettre, J., and McDonnell, R. (2020). The effect of gender and attractiveness of motion on proximity in virtual reality. *ACM Trans. Appl. Percept.*, 17(4).

[Ziemke, 2015] Ziemke, T. (2015). What's that thing called embodiment? *Avant: Trends in Interdisciplinary Studies*, 6(2-3):161–174.