EXPLORING AVATAR FACIAL FIDELITY AND EMOTIONAL EXPRESSIONS ON OBSERVER PERCEPTIONS OF THE UNCANNY VALLEY

An Undergradute Honours Thesis

in the

Faculty of Science and Information Technology
School of Design, Communication and IT
University of Newcastle
Callaghan, NSW
November, 2015

by

Jacqueline Deanna Bailey

ABSTRACT

Avatars have been a traditional mainstay of game based interactive entertainment, where they aim to enhance story-based interaction and player engagement. However, there has been an increase in avatar roles in fields such as serious gaming and simulation training, where accurate and cost-effective avatar development and ability to convey human emotional expressions is of interest. The area of emotional expression in avatars is not well understood, and *uncanniness* in avatars can pose issues that may impact on training outcomes.

There are two aims of this research, firstly to explore how avatar fidelity or realism influences the emotional experience of interactions between humans and computer-generated avatars. Secondly, to examine how the emotional expressions displayed by avatar facial features affect participants perceived valence or the intrinsic attractiveness of the avatar. In order to test these affects, this research uses a combination of survey and experimental methodologies. Utilising a Godspeed survey to measure the perception of an avatars '*humanness'*, '*eeriness'* and *'attractiveness',* and a three-part experiment measuring participant startle reflex responses to differing fidelity and emotional expression avatars, human-avatar interaction was explored.

The analysis of results indicated that participant gender played a role in the perception of avatars. In addition, the avatars themselves appear to have a significant impact on the responses from participants. The emotional expressions displayed indicated that sad expressions are less unpleasant and possibly less uncanny, than smiling.

In conclusion, this research represents an entry point into a broad, cross-disciplinary research area. While there are important findings and contributions made, the significant amount of data generated through the experiments will pose questions for future work in this research area.

ACKNOWLEDGEMENTS

Firstly, I would like to express my genuine thanks to my Research Supervisor Dr Karen Blackmore for all her guidance, advice and unfailing patience throughout this research project and thesis production. I also wish to give thanks to Megan Valentine from the Faculty of Science and Information Technology Statistical Support Service for her assistance with statistical analysis found in this thesis. I also would like to express my gratitude to the staff from the University of Southern California's (USC) Institute for Technologies Graphics Lab who supplied assistance through identifiying the high-fidelity avatars utilised in the experiments of this research. My sincere thanks also to Grant Robison and Andrew Guy from the Australian Defence College Simulation Centre for all the information and support they gave throughout this research. Finally, I would like to thank my family and friends for their encouragement and support for the duration of this research project.

TABLE OF CONTENTS

1 INTRODUCTION....1

1.1 Problem Definition....1

1.2 Research Questions....3

1.3 Approach and Contributions....4

1.4 Organization of this Thesis....5

2 LITERATURE REVIEW....6

2.1 Systematic Literature Review....6

2.2 Formulation of Research Questions....8

2.3 Locating Relevant Literature....9

2.4 Selection and Evaluation of Literature....11

2.5 Analysis and Synthesis of Systematic Literature Review....12

2.5.1 Avatar and Human Interaction....12

2.5.2 Facial Animation of Avatars....17

2.5.3 The Study of Emotional Expression in Avatars....20

2.5.4 Avatars and Levels of Fidelity/Realism....23

2.6 The Uncanny Valley....27

2.7 Review of Literature Addressing Measurement of Emotional Responses....33

2.7.1 Emotional Responses and Affective Processing....33

2.7.2 Survey Instruments (Godspeed Index)....35

2.7.3 Physiological Measurement of Emotions (Startle Reflex)....37

2.7.4 International Affective Picture System....40

2.8 Summary....41

3 METHODS....45

3.1 Introduction 45
3.2 Research Questions 45
3.3 Research Framework 46
3.4 Justification of the Methodology 49
3.5 Research Methods 52
3.5.1 Participant numbers and recruitment 54
3.5.2 Experiment 55
3.5.3 Surveys 62
3.5.4 Collection and Storage of Data 63
3.5.5 Analysis of Godspeed Survey Data 64
3.5.6 EMG Startle Reflex Data Analysis 65
3.6 Limitations of Research Methodology 66
3.6.1 Time restrictions 66
3.6.2 Participant Numbers 67
3.6.3 Limitations of choice relating to physiological indicators and avatar generation 67
3.7 Summary 68
4 RESULTS 70
4.1 Introduction 70
4.2 Participant Demographic Results 70
4.3 EMG Startle Results 72
4.3.1 EMG Startle Responses to the Baseline Images 72
4.3.2 EMG Startle Responses to the Emotional Expression Avatars 77
4.4 Godspeed Results 81
4.4.1 Repeated Measures ANOVA for Fidelity avatars 81
4.4.2 Repeated Measures ANOVA of Still and Animated Avatars 83

4.5 Other Findings – Gender Differences 85
4.5.1 EMG Data: *Baseline* 85
4.5.2 EMG Responses to Fidelity Avatars – Gender Differences 86
4.5.3 Godspeed Results – Gender Differences 88
4.6 Correlation Analysis of EMG and Godspeed Data 89
4.6.1 Fidelity Avatars 89
4.7 Emotional Expression Avatars – Still Images 90
4.8 Emotional Expression Avatars – Animated 90
4.9 Summary of Results 91
5 DISCUSSION AND CONCLUSIONS 93
5.1 Introduction 93
5.2 Contributions of the Research 96
5.2.1 Contributions from the Literature Review 96
5.2.2 Contributions from the Experiments 97
5.3 Limitations and Future Work 102
5.4 Conclusion 103
6 REFERENCES 104
7 APPENDIX A – EXPERIMENT PRE-QUESTIONNAIRE 112
8 APPENDIX B – EXPERIMENT GODSPEED SURVEY 115
9 APPENDIX C – RISK ASSESSMENT FOR THIS RESEARCH 116
9.1 Risk Identification 116
9.2 Risk Assessment – Scale 1 low 3 Medium 5 High 118
9.3 Response Development 119
10 APPENDIX D – Gender differences in godspeed means 122
11 APPENDIX E – Correlation Matrixes for all avatars 124

11.1 High-Fidelity Avatars 124
11.2 Low-Fidelity Avatars 125
11.3 Emotional Expression Avatars – Still Images 126
11.4 Emotional Expression Avatars – Animated 127

LIST OF FIGURES

Figure 1 – The Uncanny Valley (Mori, 1970) 3

Figure 2 – Systematic Literature review process 8

Figure 3 – Graphical representation of the valence/affect model (Valenza et al., 2011) 34

Figure 4 – Organisation of the startle reflex eyeblink and physiological indicators (Witvliet & Vrana, 1995, pg. 441) 38

Figure 5 – Experiment Procedure 53

Figure 6 - EMG Electrode placement on participants 56

Figure 7 – Sample of images from selected categories from the IAPS 57

Figure 8 – Age ranges of all participants 71

Figure 9 – Number of hours a week spent engaging with avatars for all participants.....72

Figure 10 – All participant responses to the baseline image categories 73

Figure 11 – All participant responses to the baseline image categories, excluding P0013 73

Figure 12 – A comparison of male and female participant response to all three categories of baseline imagery 74

Figure 13 – Average normalised peak EMG for all participants viewing fidelity avatars 75

Figure 14 - Average Normalised Peak EMG for Emotionally Expressive avatars – still images 77

Figure 15 - Average Normalised Peak EMG for Emotionally Expressive avatars – Animated imagery 78

LIST OF TABLES

Table 1 - Parameters and search sources 9

Table 2 - Inclusion and exclusion criterial for searches 10

Table 3 – Results from the systematic review 11

Table 4 - Key dimensions of uncanniness adapted from MacDorman, Green, Ho and Kock (2009). 30

Table 5 – Valence and arousal ratings of images used in this study.....58

Table 6 – Avatars used in the fidelity section of the experiment.....59

Table 7 – Avatars used for the emotional expression section of the experiments.....61

Table 8 – T-Test results for EMG startle response.....74

Table 9 - T-Test results for EMG startle responses for Fidelity Avatars.....76

Table 10 - T-Test results for EMG startle responses for emotional expression avatars.....79

Table 11 - Godspeed means and standard errors for perception of humanness of fidelity avatars 82

Table 12 - Godspeed means and standard errors for perception of eeriness of fidelity avatars 82

Table 13 - Godspeed means and standard errors for perception of attractiveness of fidelity avatars 83

Table 14 - Godspeed means and standard errors for perception of humanness of still and animated avatars.....84

Table 15 - Godspeed means and standard errors for perception of eeriness of still and animated avatars.....84

Table 16 - Godspeed means and standard errors for perception of attractiveness of still and animated avatars....85
Table 17 – Means and standard deviations for female and male participants' baseline responses....85
Table 18 - Results from T-Tests based on participant gender for Fidelity avatars....86
Table 19 - Results from T-Tests based on participant gender for Emotionally Expressive avatars....87
Table 20 – Female Godspeed Means....122
Table 21 – Male Godspeed Means....123
Table 22 – Ira (Happy)....124
Table 23 – Ira (Sad)....124
Table 24 – Emily (Happy)....124
Table 25 – Emily (Sad)....124
Table 26 – Liliwen (Happy)....125
Table 27 – Liliwen (Sad)....125
Table 28 – Leo (Happy)....125
Table 29 – Leo (Sad)....125
Table 30 - Curls (Happy)....126
Table 31 – Curls (Sad)....126
Table 32 – Jacqueline (Happy)....126
Table 33 – Jacqueline (Sad)....126
Table 34 - Macaw (Happy)....126
Table 35 – Macaw (Sad)....127
Table 36 – Curls (Happy)....127
Table 37 - Curls (Sad)....127

Table 38 - Macaw (Happy) 127

Table 39 - Macaw (Sad) 127

Table 40 - Jacqueline (Happy) 128

Table 41 - Jacqueline (Sad) 128

1 INTRODUCTION

1.1 Problem Definition

Digital characters (avatars) play an important role in the experience of video games. However, avatars are also widely used in applications not exclusively for entertainment purposes, such as in 'serious games.' Serious games often make use of avatars to achieve training and learning outcomes for 'players', featuring heavily in serious games and applications in fields such as medicine, education and the military.

Avatars are employed in the medical field by using actors portraying digital versions of standardised patients. These actor-driven avatar patients strive to educate and evaluate a medical professionals' skills in regards to physical examination, communication and level of professionalism (Rosen, 2008). The educational field uses avatars as a form of learning by substituting or supporting lectures and texts with games. These games are usually designed to be visually dynamic, with rapid pacing that leads to the student being deeply engaged with the material through a highly gratifying pictorial experience (Annetta, Murray, Laird, Bohr, & Park, 2006). In addition, the military use computer-based training systems, sometimes referred to as 'lightweight simulations', that are web or PC-based systems and designed to provide individual instruction on specific mission skills (Alexander, Brunyé, Sidman, & Weil, 2005). Whether directly controlled by human actors or instructors, or driven by artificial intelligence processes, an element of self-reflection is also a key aspect of avatar interactions.

Avatars are a virtual representation of a user who is interacting with a chosen platform. The word avatar originates from the Sanskrit for 'incarnation,' has come to refer to a user's online representation of oneself (De Zwart & Lindsay, 2012). These digital or virtual representations of a person can allow for anonymity of users or be used a form creative

expression of identity (Vasalou, & Joinson, 2009). Avatars can be expressed in a variety of forms including full body, upper body and facial/head only representations.

In this research the focus is on the facial expressions of avatars. This research seeks to extend the base of knowledge that surrounds avatars and their facial movements. This is explored by evaluating how avatar *fidelity* and *emotional expression* influences the emotional experience of the interactions between humans and computer generated avatars. In addition, an evaluation was performed on how pleasant or unpleasant the facial expressions are perceived in a series of still and animated avatars and how this relates to the concept of the 'Uncanny Valley'.

The term 'Uncanny Valley' was originally applied to the field of robotics by Masahiro Mori in his 1970 essay "The Uncanny Valley". The dip in the continuum of uneasiness or uncanniness is commonly referred to as the Uncanny Valley (Figure 1); a state that occurs in observers when a human representation looks and moves almost, but not exactly, like a real human being (Mori, MacDorman & Kageki, 2012). Although originally coined to describe this effect in robots, the term has been extended into the field of digital characters, as similar uncanny effects have become apparent, leading to problems with animating digital characters.

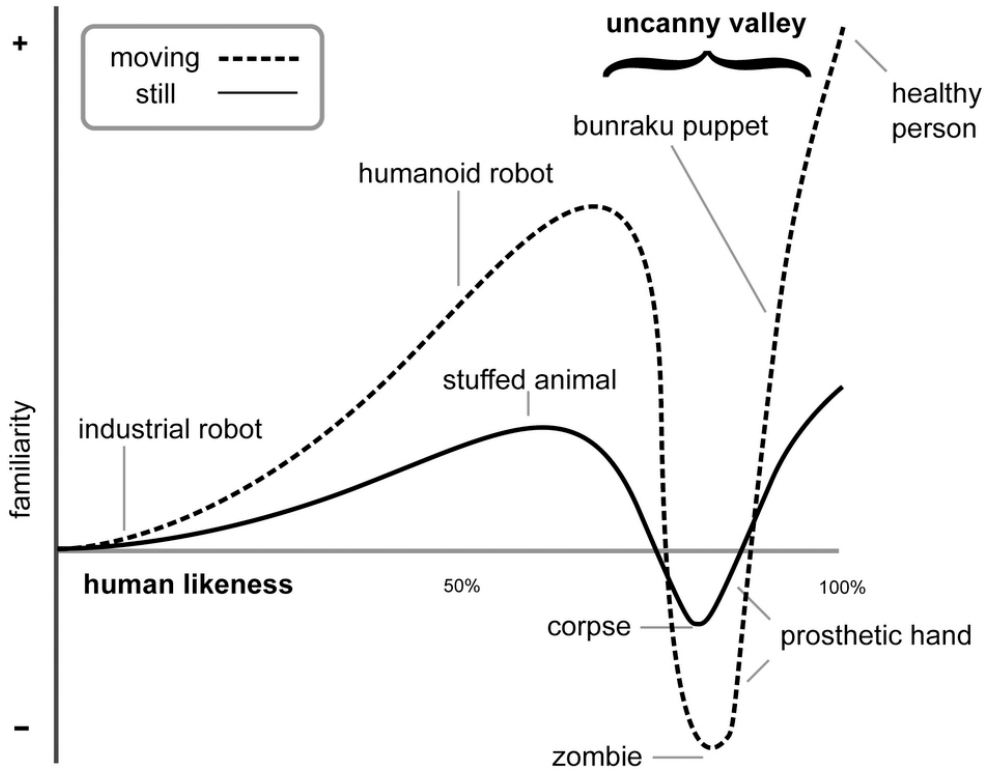


Figure 1 – The Uncanny Valley (Mori, 1970)

According to Tinwell (2014), animation of emotional expression in regards to avatars can lead to problems associated with the Uncanny Valley if done incorrectly. Avatars currently struggle to fully replicate and express emotions in a proper manner, which can lead to levels of uncanniness being experienced (Tinwell, 2014). While much research has occurred in the area of avatars, there are still several key problems that exist in the animation of avatar facial features. These include the upper facial region, the eyes, the synchronization of the mouth and the over exaggeration of the articulation in relation to speech (Tinwell, 2014).

1.2 Research Questions

In order to extend the existing knowledge of how avatar fidelity influences the emotional experience in human-computer interactions with avatars, and how facial

expressions are perceived in relation to the Uncanny Valley, this thesis aims to answer the following questions:

Research Question 1: ***How does avatar fidelity or realism influence the emotional experience of the interactions between humans and computer-generated avatars?***

Research Question 2: ***How do the emotional expressions of avatar facial features affect participants in relation to the levels of valence experienced?***

1.3 Approach and Contributions

The approach used to answer the research questions posed in this thesis combines survey instruments with physiological measures. The survey utilised for this research is a modified Godspeed survey specifically designed for avatars. This survey allowed participants to evaluate the avatars presented to them based on their perceptions of an avatars 'humanness', 'eeriness' and 'attractiveness'. In order to test how the avatars presented to them affect a participant, a physiological measurement was obtained through the measurement of the human startle reflex. This technique will allow for a quantified reaction of affect experienced by a participant when their startle reflex is engaged. From this response the data will be analysed in order to determine the level of affect experienced by participants.

In answering the research questions, the work presented in this thesis contributes to the understanding of human perceptions of avatar facial expressions. The results provide clear information on the effect of avatar fidelity and emotional expression on perceptions of uncanniness. Firstly, these effects are explored through levels of fidelity and realism that an avatar possesses. Secondly, the emotional expression displayed by an avatar was taken into consideration. In the experiments, an avatar would display a neutral to happy and neutral to sad expression in order to determine if there were any significant differences in how these

experssions were perceived. In addition, this study explores the emotional expression of avatars in still and animated imagery in order to test whether or not there are any differences in the format the avatars are presented in. The collection of demographic information from the participants in the research also provides insights into gender differences in these perceptions.

1.4 Organisation of this Thesis

The remainder of this thesis is divided into four chapters. In chapter 2, a systematic review of literature based upon the suggested process by Briner and Denyer (2012) is provided. The literature review is expanded through a narrative review of the literature relating to the most appropriate areas relevant to this research in order to answer the research questions outlined within Chapter 2.

Chapter 3 details the methodological approach adopted for this research. This chapter describes the data gathering implements utilised during the experiments of this research. These tools include a startle probe appratus to collect electromyographic (EMG) readings and a series of Godspeed surveys.

In Chapter 4, the results of the analysis on the datasets are presented. The data from the EMG readings and Godspeed are both tested independently. These datasets were then also tested for correlations. The datasets were also tested for any significant gender differences relating to gender of the avatar presented or the participant gender.

Finally, Chapter 5 provides a discussion of the main findings of the research, which includes the responses to each of the research questions. Limitations of the research are described and discussed, with opportunities for future work outlined in the conclution of this thesis.

2 LITERATURE REVIEW

The review of literature relevant to this research adopted both formal systematic and narrative approaches. In order to gain foundational understanding of the research area, a formal systematic literature review was undertaken. This review was guided by the defined research questions for this study. The key area of exploration was avatar-human interaction, including studies of emotional expression in avatars, and material relating to human perceptions of emotional expression in avatars.

Following this initial systematic review, a less structured narrative approach was taken to identify appropriate methods and any additional supporting literature. In particular, tools and techniques to assist with measuring human emotions and perceptions, including a survey tool for measuring perceptions of avatar uncanniness, and a physiological measurement instrument for exploring human emotional responses, were explored. A key technique discovered in the review was the measurement of the human startle reflex, used in conjunction with the International Affective Picture System (IAPS), for measurement of emotional response or affect.

This systematic review approach, together with the broader narrative review of methods, is presented in the following section.

2.1 Systematic Literature Review

Two methods exist for conducting a literature review: the narrative approach and the systematic approach. A benefit of performing a systematic literature review is that this process gives researchers the opportunity to gain a foundational understanding of the knowledge that is current within their chosen field (Briner & Denyer, 2012).

According to Keele (2007), a formal systematic literature review is a process of identification, evaluation and interpretation of the currently available research that is relevant to a particular research question or area of interest. A systematic literature review approach follows a formalised process to scope, locate, select and analyse literature relevant to an area of study. This is the approach implemented in this study, and in keeping with a systematic approach, it seeks to address specific research questions by evaluating and summarising the existing works related to avatar emotional expression. In terms of emotional expression, this study will specifically focus on the facial expression of emotions as opposed to the expressions of a full-bodied avatar.

A systematic approach to literature review can occur as a replacement, or in tandem to, a narrative review approach. A narrative literature review is considered the more traditional approach that relies on the knowledge and expertise of author. In situations where the author may not have a pre-existing understanding of the important areas and works in an area, a narrative approach is subject to methodological flaws that can lead to bias (Capriani & Geddes, 2003).

Tranfield, Denyer and Smart (2003) argue that systematic reviews offer a replicable, transparent and scientific process that differs from a traditional narrative review. The authors argue that this process aims to minimise any form of bias through intensive searches of literature utilising both published and unpublished works. This process also provides an audit trail of the procedures, decisions and conclusions made by the reviewer. Given that this study is an exploration of this field of study by the author, the systematic approach is an appropriate starting point for the review.

There are a number of approaches that exist for conducting a systematic literature review, which differ in the number of stages undertaken in the process. In regards to this thesis, the suggested approach from Briner and Denyer (2012) has been utilised. This

approach includes five primary stages that are common both within and outside of the field of clinical based studies. Briner and Denyer (2012) suggest that systematic reviews provide the researcher with a foundational understanding of the current knowledge of their chosen field.

The approach suggested by Briner and Denyer (2012) addresses a specified question by making usage of a transparent and explicit method to gather relevant literature. This then allows for a criterial evaluation of the studies located based on the literature that is available. Finally, this process will give the reviewer the appropriate means to draw conclusions about what is known and unknown within their chosen field at the time of their research being undertaken. The process used within this research is depicted in Figure 2 below.

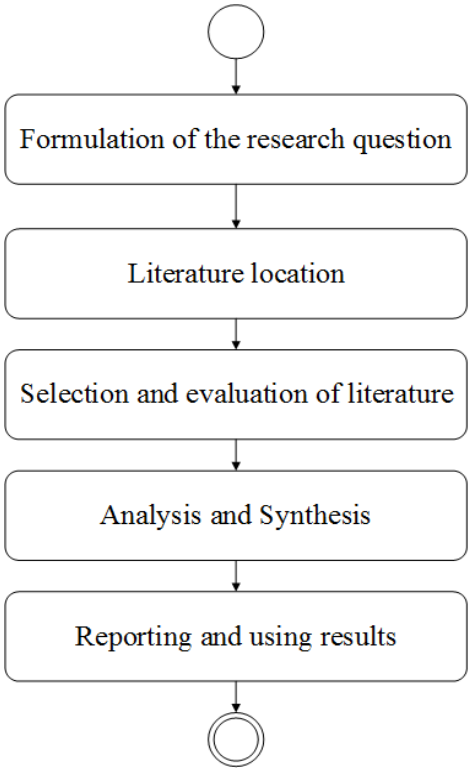


Figure 2 – Systematic Literature review process

2.2 Formulation of Research Questions

In order to identify literature that was the most relevant to this study, the following research questions were defined to guide the systematic review process:

<u>Research Question 1:</u> How does avatar fidelity or realism influence the emotional experience of the interactions between humans and computer-generated avatars?

<u>Research Question 2:</u> How do the emotional expressions of avatar facial features affect participants in relation to the levels of valence experienced?

From these questions the topic areas of avatar fidelity and realism, avatar-human interaction, emotional expression and measures and indicators of affect, characterise the literature that would be required to be explored.

2.3 Locating Relevant Literature

Relevant literature for this study, based on the above research questions, was located using a set of clearly defined search parameters. The following table (See Table 1) presents the parameters and search sources that were used for this literature review.

Table 1 - Parameters and search sources

Bibliographic Databases	Web of Science, Scopus
Search Engines	Google, Google Scholar
Article Type	Journal articles, conference papers
Search On	Title
Sorting on Return	Relevance
Language	English
Period of Publication	Unrestricted

Inclusion and exclusion criteria are used in the systematic review process to specify the specific search terms that will be used to locate relevant literature. The above parameters and search sources were configured with the following inclusion and exclusion criteria for searches relating to the literature location (Table 2).

Table 2 - Inclusion and exclusion criterial for searches

Component 1: Compulsory	“Avatar*”, “Virtual character*”
Component 2: Compulsory	“Emotion*”
Component 3: Compulsory	“Expression*”
Component 4: Compulsory	“Fidelity”, “Realism”
Component 5: Excluded	“Robot*”

While there has been much work involving the field of robotics and the Uncanny Valley this study focuses primarily on avatars. Therefore there was a need to exclude ‘robot*’ from the search criteria in order to optimise search results for this research. With the combination of the criteria specified above, the following Boolean search string was used:

> (*avatar* OR "virtual character*") AND (fidelity OR realism OR emotion* OR expression) NOT robot**

This Boolean string returned a total of 162 results that were then selected and evaluated for this review in order to gain a foundational understanding for this study. The selection and evaluation process is discussed in the following section.

2.4 Selection and Evaluation of Literature

The primary areas of relevance for this study predominantly relate to human perceptions of avatars, and the levels of emotional expression that can be generated by avatars. This is in addition to the levels of realism that can be produced from various techniques used to create the avatars. Studies identified in this section were filtered to ensure that they relate to these topics and inform the overall research questions.

In order to locate relevant literature, a title search was undertaken to identify the literature discussed below. Following this, abstracts were reviewed to determine the relevancy to this research of the returned articles, and were organised on this basis as a result of a systematic review. Where relevancy was unable to be determined from the abstract, a review of the entire article was undertaken. Table 3 provides the total numbers of articles located, in addition to those that were considered relevant. Duplicate articles were removed during this process.

Table 3 – Results from the systematic review

Database	**Found**	**Relevant**	**Duplicates**
Scopus	53	8	17
Web of Science	100	13	17
Google Scholar	9	6	3

The search results described above form the basis of the selection and evaluation of the papers found as a result of the systematic review. From these results, five specific areas of research emerged, that include:

- avatar and human interaction,
- facial animation of avatars,
- the study of emotional expression in avatars,

- avatars and levels of fidelity/realism, and
- tools for measuring physiological reaction.

The analysis and synthesis of the literature located is presented in the following sections.

2.5 Analysis and Synthesis of Systematic Literature Review

The literature reviewed in the following section focuses on prior research directly relating to, and informing, the research questions in this thesis. However, also identified were articles that while not directly relevant, provide important contextual information or highlight opportunities for future research in this topic area.

2.5.1 Avatar and Human Interaction

Literature on avatar and human interaction revealed several common themes of relevance. These include the ways in which humans and avatars interact with each other, and the possible ramifications of these interactions. Similarly, another theme relates to the use of avatars as identity for human users. Social interaction emerged as another key field in the study of the relationship between human and avatar. The literature also indicates that there are multiple uses for avatars in many fields, although the role the avatar has is often similar. Finally, the topic of avatar ethnicity and cross-cultural perceptions of avatars emerged as an important topic of study. These various themes and their relevance to the research are discussed in this section.

A key area of the literature focuses on avatar and human interaction, and specifically, how avatars affect human users. In particular, this theme relates to how avatars can affect a human users' emotional state during the interactions between humans and avatars. In their study on how the physical similarity of avatars can influence the learning of emotion

regulation strategies in teenagers, Wrzesien, Rodríguez, Rey, Alcañiz, Baños and Vara (2015) demonstrate that avatars can influence human behaviour when the avatar presented is stylistically similar to the user. Their results indicate that avatars that are similar in appearance to participants have a significant impact on the emotional valence and arousal levels of that participant. In addition, the intensity of the participants' emotional states are greater in comparison to when participants observed a neutral avatar. This suggests that valence and arousal are useful dimensions for assessing participant responses, and also that avatars that are more similar in appearance to participants, or indeed realistic, should also stimulate a more intense emotional response.

In contrast, Vaškevičius, Bazevičius, Vidugirienė and Kaminskas (2013) have adapted and developed methods that can affect the users emotional state by altering visually observed parameters within a virtual character. In their study three sets of faces were used to explore the parameters that could be used for emotional-state control. Facial features of the avatars were altered in order to gauge participants' reactions on the parameters of excitement, meditation, engagement/boredom and frustration. These reactions were monitored through electroencephalogram **(**EEG)-based equipment. This form of testing allows for detection of abnormalities related to brain activity in the form of recording brain wave patterns. The authors conclude that their results could be utilised to create a feedback model based on a psycho-emotional state control model within virtual reality environments. Similarly, the experiments undertaken in their study lead to [a] user feedback based model that facilitates the control of the emotional-state of a user via an avatar (Vaškevičius et al., 2013).

These studies demonstrate that interactions between humans and avatars can have an affect on the emotional state of the human. The method of developing feedback techniques that would aid in affecting the emotional state of a user, as produced Vaškevičius et al.,

(2013), demonstrate that experimental approaches have been developed that specifically map avatar expression to the desired emotional response in human users.

Another theme that was observed from the literature was the use of avatars as identity for users. As this research is investigating the emotional experience of human-avatar interaction it is important to consider that self-expression through these avatars could be seen as a form of emotional expression. For example, people using avatars as a form of identity within social interaction communities can also become a part of an artificial society with its own customs and beliefs. This could potentially enhance their understanding of avatars and their expressions in order to stimulate the users own emotional responses or to convey emotional experiences to other users.

The use of avatars can take many forms, and for a variety of reasons, however it is broadly recognised that using an avatar as an identity for a user can be seen as a form of self-expression. Harrell and Harrell (2012) argue that in order for people to participate within virtual worlds as players or learners, they need to be able to build a computational self-representation of themselves. This can enable proxies for people in order to interact with communities. The literature also indicates that this type of identity concealment is not limited to these areas; online chat systems also allow for this sort of interaction through avatars. This area was explored by Liu, Chen, Gao, Xie and Gao, in their 2008 article, "AAML Based Avatar Animation with Personalized Expression for Online Chatting System." In this research, the authors proposed a prototype system for chatting online supported by XML based facial animation language. Liu, Chen, Gao, Xie and Gao (2008) used exaggerated caricatures and facial animation to allow people to freely express their emotions to others they were communicating with. However, their work focussed on the process for developing the avatars rather than evaluating the utility of the approach for

developing perceivable avatar emotional expressions. Despite this, the results raise questions regarding the role of avatar realism in facilitating communication and interaction.

Social interaction between users through the use of avatars can lead to the formation of communities. For example, Liu (2006) suggests that 3D artificial societies can be constructed within virtual environments for virtual charters to interact with emotionally. Within this society, new and existing social norms are created and maintained, allowing for a new level of interaction between people in these communities. Importantly for this research, Liu (2006) also argues that areas such as culture differences, emotion, motivation and personality are factors of social interaction that can be integrated into virtual characters. However, the use of avatars as a form of socialisation is only one of the uses of avatars, with more serious purposes of increasing prominence.

The use avatars for simulation and training, in fields including industrial environments and education, are a growing area. According to Ziegeler and Zuehlke (2005), the use of avatars in industry environments is an emerging field that requires extensive time and effort to move it into industrial application. In their study, a set of characters was designed to aid users in the easy recognition of text based screen messages. Avatars used set facial expressions and a relevant uniform to assist in communicating different types of information via the text. For example, a smiling avatar dressed in overalls was used to communicate maintenance information whereas an avatar with an alarmed expression, dressed in red military uniform, was used to communicate alarm messages. Ziegeler and Zuehlke (2005) argue that parts of the systems involving manufacturing and processing are becoming more complex with a huge amount of information needing to be visualised on the user interface (UI). This could potentially add an element of enjoyment to the communication of information that may motivate the end user through support, and produce a positive

working environment. Importantly, they highlight that the style and expression of an avatar is also crucial to improving information transfer.

Similarly in education, Guynup, Broglio and Demmers (2004) argue that a major disadvantage of online learning is the lack of a human dimension. That is, an absence of the sense that the users are a part of a larger community of learners. However, with the aid of a virtual space and avatars, the educator transforms into an agent of new learning via a shape-shifting entity. This can facilitate engagement with materials and other learners seamlessly in the digital space. The use of avatars as a tool for learning is present in both these examples, and they highlight the role of avatars in bridging the real and virtual worlds of participants and learners. However, they also raise issues regarding the style of avatar used in such situations.

Finally, within the avatar and human interaction theme, the literature revealed that there are a number of issues in regards to avatar ethnicity and cross-cultural expectations. While the use of avatars is growing within society, it appears that the issue of ethnicity in the creation of avatars is addressed only minimally. For example, after examining 60 online and 20 offline massively multiplayer role playing games (MMRPGs), Dietrich (2013) found a significant lack of ability to customise for races other than ‘white’. In general, the vast majority of games located in this study do not allow for the creation of avatars with a non-white racial appearance. Where non-white skin colour for avatars is an option, a further complicating factor is the lack of realistic non-caucasian facial structures. While Dietrich claims these issues lead to the potential divide between racial and ethnic minorities being reinforced in gaming, the implications of this in relation to the broader issue of user perceptions was not considered.

In another study related to cross-culture expectations of avatars, Koda and Ishida (2006) explore the cultural differences in interpreting avatar facial expressions. They argue

that avatars must be designed carefully in order to convey the most accurate examples of emotional expression regardless of culture, even though misinterpretation is less likely between positive and negative expressions. Avatar ethnicity and cross-culture expectations need to be considered in regards to the potential global audience that will utilise these avatars, in order to accommodate a range of people and their cultures.

These studies raise interesting and important points around the impact of cultural diversity in the both the communication of emotional expression in avatars, and also in the ways that individuals from different cultural backgrounds may perceive avatars. In light of this, some consideration to incorporating cultural diversity into this research is adopted, however given the complexity of the issues, it is not a focus of the research. In the following section of this document, research considering approaches to facial animation of avatars will be discussed.

2.5.2 Facial Animation of Avatars

The systematic review of literature revealed that there is currently no standard method or approach to animating avatar facial features. The readings indicate that there are several prominent techniques, with widely varied approaches to facial animation. However, there has been no formal measure of success in regards to what makes facial animation generation successful, or indeed how to measure success. In addition, a number of additional areas of interest were located within this theme. These include discrete facial expressions and the reproduction of eyes in avatars. These techniques and considerations will be discussed in this section.

As there is no universal method for faithful recreation of facial animation in avatars identified, this section will discuss several key techniques found in the literature. The approaches to animation of avatar facial features reflect the complex nature of faithful

reproduction, and the examples given here provide an illustration of the difficulty of this task, and consequential lack of a universal solution. An exhaustive listing of techniques utilised for animation of avatar facial features is beyond the scope of this study.

Zalewski and Gong (2005) address the issue of modelling facial expressions of avatars in their study of 2D statistical models of facial expressions for creating realistic 3D avatar animations. They demonstrate that their approach, which uses relatively simple models, can efficiently produce a convincing level of realism. Their technique used hierarchical facial decomposition to capture the manifolds, or topological spaces, of human facial expressions. However, their assessment of realism was based on a single item scale evaluated by 16 human test subjects limiting the generalisability of results.

In contrast to the relatively simple avatar generation process used by Zalewski and Gong (2005), Puklavage, Pirela, Gonzalez and Georgiopoulos (2010) used a more detailed approach involving a machine learning method. Specifically, they employed reproduced personalised facial expressions in a computer avatar through a system from the Facial Action Coding System which is mapped by a Particle Swarm Optimisation algorithm. This was then transferred onto a computer-generated rig that had also been based on Facial Action Coding System. Evaluation of their approach focused on the efficiency of the development process, with no human evaluation of the ability for the generated avatars to realistically convey emotional expressions occurred.

Another technique for avatar emotional expression generation includes an efficient economic video-driven creation process as proposed by Chiang, Chen, and Yang (2013). This method allows for instantaneous creation of a huge variety of virtual characters and automatic syntheses of facial animation. The method employed was a component-based approach, which allows for person-to-person expression transfers as well as, component-based creation

for novel avatars. Again this work focuses on the value of the creation approach rather than the communicative effectiveness of the avatars produced.

Similarly, Seddik, Maâmatou, Gazzah, Chateau, and Ben Amara (2013) propose a solution capable of recognising facial expressions performed by a persons' face, and mapping them to a 3D face virtual model using a Microsoft Kinect. The approach developed has the advantage of being unsupervised, and the mapping between inputs and pre-constructed emotional expressions produces reliable results. This approach also has the benefit of allowing for real time capture and transfer of emotional expressions from a human actor to an avatar, which is useful for education and training scenarios. However, reliability was assessed via the ability of the process to correctly map inputs to the pre-constructed expressions rather than human assessment of the expressions produced.

Given the broad range of potential applications, this research will make use of a similar technique to that of Seddik et al. (2013) by utilising Faceshift (Version 2014.2.01, Faceshift AG, 2015) to create avatars for the experiments. Faceshift is a markerless motion-capture software that allows for the tracking of facial expressions via a sensor in real-time (Weise, Bouaziz, Li & Pauly, 2011). This tool will allow for facial expressions to be performed by the researcher in order to generate emotional expressions on avatars for the experiments in this study with ease. Simplifying the creation of the avatars will correctly direct efforts to the assessment of the feasibility of the produced expressions to convey emotions.

Facial expressions are considered to be one of the most expressive ways for human beings to convey their emotion, beyond verbally expressing them in face-to-face communications (Moser, Derntl, Robinson, Fink, Gur, & Grammer, 2007). The concept of minor or discrete facial animations have a role in assisting communication in the digital space. For example, one technique that can enhance the believability of an avatars emotional

expression is through the inclusion of spontaneous facial expressions. In their 2012 study, Yang and Bhanu explored subtle forms of expressions through a system that makes minor alterations in the overall facial expression. A single face could be segmented, based on one or more expression labels, in order to create a more believable facial expression presented by the avatar. These segments could reduce the level of uncanniness that can be associated with the lack of proper emotional expression.

Another identified feature of avatar facial expression that is important when attempting to reproduce realistic facial expressions is the eyes. Wang and Geiger (2011) contribute that eye movements can be used as emotional signals. Specifically, they hypothesise that eye movements are a crucial component for expressing emotions in communication within virtual worlds. In support of this, Steptoe and Steed (2008) identified that socket-deformation had a highly positive impact on the perceived authenticity of their avatars' overall appearance. Socket-deformation is achieved by deforming the areas of the avatars' eyelids, eyebrows and the immediate area that surrounds the eye depending on the current eye rotation. Their findings indicated that the facial expression of avatars can be enhanced with the appropriate level of eye reproduction. As well as having an impact on the capacity for avatars to deliver effective emotional expressions, small expressions may also produce avatar *uncanniness*, which is of direct relevance to the thesis and is discussed in Section 2.6.

2.5.3 The Study of Emotional Expression in Avatars

Several areas relating to the study of emotional expression in avatars will be examined in this section. These include emotional interaction between humans in virtual environments using avatars, and effects of using avatars as a form of expression in

communication. In addition, the recognition of expressions displayed by avatars, and how these emotional expressions are generated, will be explored.

One theme that emerged from the literature was the use of avatars as a form of expression in communication. The literature revealed that many areas associated with virtual environments were working to generate avatars that were emotionally expressive. Farbi and Moore (2005) argue that in order to provoke empathy amongst users of virtual environments, emotionally expressive avatars should be utilised. This could potentially allow for a greater level of interaction between users through avatars that are emotionally expressive. Importantly, these benefits only apply so long as the participants in these interactions can recognise the emotions being displayed by the avatar presented to them.

The ability for users to recognise the emotion expression generated by the avatars presented to them emerged as a key theme in the literature. In order for *effective* and *affective* communication to occur between users and avatars, users must be able to recognise the emotional expression being presented to them. Participants of a study undertaken by Noël, Dumoulin and Lindgaard (2009) correctly identified the expected emotional state of a face presented to them regardless of the level of realism of the avatar face. This is an important finding in the context of this research, as it implies that communication of emotions in virtual settings is not dependent on the human-likeness, or realism, of the avatar used to convey them.

Broadly, the literature located through the systematic review reflects that the majority of participants of studies who viewed emotional expressions on avatar faces could identify the correct emotion displayed to them. However, this was not always the case, with some counter results attributed to flaws that were found in experiments and indicated by the authors of that research. For example Noël, Dumoulin and Lindgaard (2009) discovered issues with displaying a neutral facial expressions. The authors conclude that the difficulty

experienced with the neutral face used was due to the particular avatar used within their experiment rather than a broader recognition issue.

Similarly, Noël, Dumoulin, Whalen and Stewart (2006) identified that recognition rates dropped significantly during their second experiment. The authors argue that participants are sensitive to small changes in the presentation of avatar faces, as evidenced by the changes in the stimuli utilised for their experiments. These results indicate that it is important that there is a clear indication of what emotion an avatar is trying to express in order for effective and affective communication to occur for both the participant and avatar designer. The creation of these expressions needs to avoid subtlety, and occur at a level in which a range of users can identify the emotional expression displayed to them.

As with avatar creation studies, the review of the literature also revealed that there was no standard way of generating emotional expression in avatars. For example, Kim (2007) describes a method of controlling the expressions of avatars in real-time through expressions found in the facial expression space technology of that study. The system holds approximately 2400 facial expression frames with the transition between expressions inferred from captured facial data. In contrast, Arellano, Varona and Perales (2008) use an affective model to generate avatar facial expressions. In this approach, eight higher level emotional states such as exuberant, bored and hostile, are mapped to facial definition parameters, which are defined by the location of feature points. These feature points are used to convert a given emotional state to a visual representation of this state on an avatar. Similarly Li, Liu, Zhang, Pan and Song (2010) use a newly improved three levelled affective model known as personality-emotion-mood, which could aid in generating intelligent and emotional avatars.

As there is no universal way of generating the required emotional expressions for avatars, this area remains an important area of research, with only representative approaches provided here. Importantly though, these existing approaches have succeeded in generating

recognisable emotional expressions in avatars (Noël, Dumoulin, & Lindgaard, 2009). As described in Section 2.5.2, this research uses the Faceshift markerless motion capture software to generate avatars. Within this environment, the resolution of mapping between the emotional expressions of the human actor and the associated avatar blendshapes is sufficient to generate both still and animated avatars with non-subtle emotional expressions.

Lee, Carlson, Jones, Johnson, Leigh, and Renambot, (2010) argue that humans share the unique ability to express and recognise emotions within face-to-face communication. This is an important point, as the need for effective communication to occur between humans and human-like avatars there should to be a reliable level of emotional expression and recognition from both sides of this interaction. In their paper 'Designing an expressive avatar of a real person', the authors analyse the effectiveness of emotional expressiveness of an avatar. The results of their experiments indicate that the only happy and sad expressions gained the high success rates in terms of user recognition, whereas anger, fear, surprise and disgust did not attain high recognition rates. The authors conclude that is an important finding as this indicates that the avatar created is capable of displaying happy and sad roughly to the same degree as a human counterpart. This finding is useful to the research being undertaken within this study, as it provides a framework of which emotional expressions would be best suited to study within the given time frame available in order to achieve meaningful results. In the following section of this literature review, a discussion of previous work on avatars and considerations relating to levels of fidelity and realism is presented.

2.5.4 Avatars and Levels of Fidelity/Realism

In this research, the term *fidelity* refers to the degree of exactness with which something is copied or reproduced ("Fidelity," 2015), whereas *realism* is considered to mean the quality or fact of representing a person or thing in a way that is accurate and true to life

("Realism", 2015). These definitions follow the standard for these terms from the Online Oxford Dictionary.

While not formally defined within the relevant literature, these terms, and their meanings that have been adopted for this research, are generally used in these contexts within relevant studies. Several areas of interest in regards to levels of fidelity and realism in relation to avatars were located from the systematic review. These issues include security in telecommunications, identification and emotional perception amongst users in a global audience, the usage of wrinkles as a design feature in expressive avatar facial appearances, as well as the use of colour in facial features. All of these areas contribute to the levels of realism that avatars can obtain and are thus of interest to this research.

With advancement in technology, telecommunications are becoming an alternative form of communication for businesses. According to Riek and Watson (2010), organisations are employing approaches such as telepresence to facilitate global operations. In particular, an evolving area involves employing technology to bring people together in virtual environments to meet and conduct business. However, this could lead to security vulnerabilities such as the communication channel being exposed to attack.

Based on these security issues, it is worth considering the risks of displaying high-fidelity avatars to users. This could be seen as a greater issue in relation to a potentially global audience that could be affected by these issues. For example, the use of an avatar in communication systems masks the real identity of those responsible for the words and/or actions of the avatar, opening the door to a range of trust and security issues. While important to the broader study of avatar use for communication, trust issues associated with avatars are beyond the scope of this research.

Although higher quality avatars may intuitively be linked to a better user experience, a higher the level of realism does not guarantee a greater level of users' acceptance. In their

study of the impact of avatar realism and eye gaze control on the perceived quality of communication in a shared immersive virtual environment, Garau, Slater, Vinayagamoorthy, Brogni, Steed and Sasse's (2003) study confirms findings from previous immersive studies that utilise semi-photorealistic avatars where inferred gaze significantly outperformed random gaze. The authors state that this finding allows for an inexpensive way of improving the level of expressiveness that an avatar can possess, using the benefit of inferred eye animations, to have a significant positive effect on participants' responses. This is an important finding as this study demonstrates that there are simple and low-cost techniques that can enhance the human-avatar interaction to be a more positive emotional experience for the user.

The authors also found that there is a significant interaction effect between appearance and behaviour. For example, the lower-realism avatar did not benefit from an inferred gaze, and as a result, a certain level of realism must be implemented in order for the gaze to be effective. Garau et al. go on to discuss the importance of aligning visual and behavioral realism for increased avatar effectiveness, finding that the lower-realism avatar was not outperformed by a higher level of realism. This may indicate that the higher the realism of an avatar, the greater expectation there is for realistic behaviors. Finer level facial features have also been found to impact on the perception of emotion expression in avatars.

Research by Courgeon, Buisine and Martin (2009) explores the impact of expressive wrinkles on avatars. Their research has been enabled by advances in technology that now affordably produce facial animation that has reached a high level of photorealism. The results of the study indicate that realistic wrinkles increase agents' expressivity. However, the wrinkles did not aid in the identification of the emotion an avatar was expressing. The use of wrinkles as a level of expression in regards to realism could be seen as another form of discrete expression from avatars in terms of avatar and human communication. Thus we

expect highly realistic facial movements seen in photorealistic avatars to impact on the overall user acceptance of the avatar, but to not add to the ability for the avatar to communicate emotions. The final aspect of fidelity to be explored in this section relates to the colouring of avatars faces in relation to emotional expression.

The colouring of the facial features of avatars relates to changing skin tone, and has been found to impact on perception and recognition of emotional expressions. Alkawaz and Basori (2012) argue that the colouring of the skin on the face is a parameter that affects the realism of the facial expressions generated. This is primarily because it is closely related to the emotion that is occurring within the human being, for example, reddening of the face with anger. It is thus noted that being able to produce realistic colouring of avatar faces would add another of level to the fidelity in an avatar, which could allow for a greater acceptance of a computer generated avatar. While important, this feature is noted for future studies, and is not considered in this thesis.

The previous sections provided a discussion of the key themes that emerged from the systematic review of literature relating to the facial expression of emotions in avatars, and the issues associated with avatar fidelity. The primary interest in this research is how users perceive avatars. In particular, measures of the emotional response of users to different avatar facial expressions, and levels of fidelity, are critical information requirements to address the research questions. In the following section (2.6), a review of literature relating to the specific topic of the Uncanny Valley is presented. Following this, the results of a narrative review of relevant literature on approaches to the measurement of emotional responses of human subjects to visual stimuli are provided.

2.6 The Uncanny Valley

The concept of the Uncanny Valley, was briefly introduced in Section 1.1 of this thesis. This term is used to describe what Mori (1970) theorised as a persons' response to a human-like robot would be. He described the emotional experience as moving swiftly from empathy to revulsion due to the failed attempts to maintain a life-like appearance by a human-like entity. This highlights a descent into 'eeriness' known as the Uncanny Valley. Given the general perception that uncanniness is an unpleasant feature, valence emerges as the key emotional measure for this research.

The previous discussion of literature relating to avatars, and levels of realism and fidelity, raise some salient points relating to the Uncanny Valley. The level of fidelity and realism associated with an avatar can lead to various issues relating to the lack of faithful reproduction of human-like entities. While these human-like characters may have a similar appearance and movements, the lack of complete fidelity and realism can lead to the feeling of eeriness associated with human-avatar interaction. If the lack of fidelity and realism is high, a more unpleasant emotional experience could occur for the user. For example, if a low-fidelity avatar matches their appearance with less realistic animation of their eyes or mouth the level of uncanniness would not likely increase. Whereas, in an avatar with high realism there would be an expectation that the avatar would perform as more human-like with more realistic animation in place to accommodate the eye and mouth movements. However, this is often not the case and the level of uncanniness increases in the user when viewing this avatar. It is reasonable to consider that a higher the level of unpleasantness could be experienced by a user when there is a higher level of uncanniness. Therefore this study makes use of measuring emotional responses in order to gauge the reactions of participants in relation to the avatars they are viewing.

The concept of uncanniness also requires discussion in order to clarify how this relates to the Uncanny Valley and the research being undertaken in this work. Firstly, there is a consideration of what is meant by the term 'uncanny', how it is defined and what the features of uncanniness are. Secondly, how people respond to something that is uncanny will be explored in terms of how people react to the feeling of uncanniness. Finally, the issue of uncanniness and its relation to the level of realism and 'attractiveness' of an avatar will be examined.

There have been two key developments in the study of uncanniness in the literature. Early work focuses on uncanniness from perspective of film and story-telling and is based on the work of Jentsch (1906) and Freud (1919), aligning uncanny with unfamiliar. Proulx, Heine and Vohs (2010) consider uncanniness as the feeling that is aroused by unfamiliar experiences that occur within familiar situations. The authors argue that within completely unfamiliar settings, expectations for normality and predictability are lower, and therefore there are fewer events that would disrupt an individual's consciousness and trigger a sense of uncanniness. Within these settings there may be a higher level of expectation to experience the sense of unexpectedness. This argument builds upon Freud's work, The Uncanny (1919), in which it is stated that only the unfamiliar familiar (*unheimliche heimliche*) threatens established meanings and the feeling of uncanniness is aroused.

Similarly, Jentsch (1906) argues that the effect of the uncanny can be achieved easily when an attempt to reinterpret a lifeless entity into a poetic or fantastic way. This is especially true in anthropomorphic terms, due the natural tendency for human beings to presume that these entities are animated in the same manner as themselves. Furthermore, Jentsch indicates that one of the predominate causes of uncanniness is doubt as to whether or not an entity presented is a lifeless object or is animated. This emotional experience lasts until the doubts of the viewer are resolved and a new emotional state takes its place. This is

important concept in terms of this research, as the word uncanny is described by Jentsch as a lack of orientation that correlates to the uncanniness of a thing. If an entity such as an avatar is unpleasant, the viewer deems the avatar to be a lifeless entity and responds negatively to the avatar itself. It is relevant that this work makes use of measuring perceived pleasantness of the avatars in regards to making the viewer feel at ease with the avatar presented to them.

However, the uncanny can be used to deliberately provoke feelings of eeriness and horror in order to arouse and excite viewers. For example, Brenton, Gillies, Ballin and Chatting (2005) argue that movie directors have exploited audiences for uncanny reactions, in order to heighten the sense of fear and dread experienced by the audience. In their essay, "The Uncanny Valley: does it exist?" the example provided is Frankenstein's monster. This creature lives in the uncertain realm between the living and the dead provoking the arousal of eeriness and horror amongst the audience. The affect on the emotional experience of the viewer results in the emotional state of being both disgusted and sympathetic towards to entity presented to them.

Uncanniness can be related to a thing or incident that is outside the normal expectations of human beings, eeriness can be linked to negative emotion experiences. As discussed by Burleigh, Schoenherr and Lacroix (2013), this eeriness can be applied to digital replications of humans. Their essay included an example from the movie "The Polar Express" (Goetzman, Starkey, Teitler & Zemeckis, 2004), where digital characters were modeled and animated based upon the movements and features of the actors portraying them. Due to the high level of realism many authors have suggested that the near-perfect likeness of a human being was the main cause of the negative emotional experiences of the audience. Evident from this work is that viewing material considered to be uncanny produces an emotional response.

Tinwell (2014) quotes both Jentsch (1906) and Freud (1919) in regards to the emotional experience of the uncanny. Jentsch described the uncanny as a human beings cognitive dissonance of being unable to decide if a thing presented to them is real/unreal or alive/dead. Whereas, Freud characterised the uncanny as the familiar behaving in a manner that is unfamiliar raising alarm in the viewer, in which the object becomes regarded as repulsive or abhorrent. These concepts are later revisited by Masahiro Mori in his 1970 essay *'The Uncanny Valley'* in which he developed a hypothetical graph representing the uncanny valley.

How people respond to the feeling of uncanniness varies. Kierkegaard (1843) argued that uncanniness was not associated with thrills like freedom, but with emotions such as fear and trembling. MacDorman, Green, Ho and Koch (2009) suggest that people react to uncanniness as threat avoidance, shared circuits for empathy and evolutionary aesthetics. These ideas are important, and suggest that using emotional experience of viewers is a valid way to measure the uncanniness of an entity presented to them. These concepts can relate to the key dimensions of uncanniness as discussed in the Table 4 below.

Table 4 - Key dimensions of uncanniness adapted from MacDorman, Green, Ho and Kock (2009).

Key Dimension	Description	Relevance to Project
Specialised Perceptual Processing		
Threat Avoidance	Related to the human need of self-preservation, hence perceived defects in a virtual human entity may trigger a negative response	Any perceived defects in the avatars presented, potentially represent a threat to the viewer making the interaction an emotionally unpleasant experience.

Shared circuits for empathy	The perceptual, cognitive and affective processing working together in order to perceive uncanny forms	Human beings share emotional experiences in order to understand others intentions, even though avatars possess human appearance and emotional expressiveness, their forms are perceived as uncanny. Human beings tend to empathise with the avatar by imagining themselves in the avatars place. Which makes the viewers' interaction an emotionally unpleasant experience.
Evolutionary aesthetics	The perception of attractiveness as a biological basis used within a specialised perceptual processing which is both automatic and stimulus derived	Human beings have evolved to perceive attractiveness in order to locate a potential mate. If the attractiveness of an avatar is low it can lead to the viewer having a negative reaction to the avatar presented to them.
Cognitive Processing		
Cognitive dissonance of liminal objects	Refers evolutionary aesthetic norms whereby humans did not evolve with robots or animated characters which may result in perceived challenged to their makers uniqueness	When presented with an entity that looks and acts similarly to themselves human beings tend to experience an unpleasant emotion experience. As, avatars may represent a threat to the human beings core identity causing a negative reaction from within them.

Sorites paradoxes involving personal and human identity	Relates to an argument over when exactly does life begin, and whether or not we are all just machines, machines with no hope for a continuation after our death	Avatars that are highly similar to human beings tend to raise a challenge of how human development and mortality is defined. If these definitions do not correlate to existing accepted norms, the viewer tends to experience a heightened sense of uncanniness related to the avatar in the face of their own mortality.
Terror Management theory	Demonstrates how subliminal reminders of our own mortality can cause a shift in our attitudes and preferences. In addition to our personal identity is socially constructed from our cultural world view, therefore linking the self to a higher concept provides solace in the face of our mortality. The presence of animated characters challenges this uniqueness of humanity and undermines the sense of human and personal identity.	If the worldview of an observer is disrupted due to the presence of avatars, this can cause a sudden lack of identity. The viewer would experience an unpleasant emotional response to any avatar that might cause them to question their identity, uniqueness and mortality.

Finally, the level of realism and attractiveness of the stimulus presented to a viewer could potentially induce a heightened sense of uncanniness. Brenton, Gillies, Ballin and Chatting (2005) state that by increasing the level of realism in the stimuli presented, the level of sensitivity to cues that would indicate falsehood become greater. This idea argues that the Uncanny Valley is related to the human ability to extract emotional and social information from a human face that is presented to them is important. This suggests that utilising a way to measure unpleasantness is a valid way to study avatar-human interaction, as the viewer has no difficulty in identifying the variations of facial expression.

In addition, Seyama and Nagayama (2007) argue that in everyday life people do not generally confuse artificial and human faces. This is also an important idea, suggesting that the human visual system has a high level of sensitivity to the realism of a face that is displayed to them. This could indicate that humans are hypersensitive to uncanniness, which enacts the emotional experience of unpleasantness in the presence of the uncanny. Seyama, and Nagayama, (2007) conclude that in order to understand the Uncanny Valley's effect, consideration must be given to both the level of realism and abnormality of the virtual human presented to the viewers.

2.7 Review of Literature Addressing Measurement of Emotional Responses

A number of tools are available for measuring the emotional responses of participants. Generally, measurement of emotions is considered in terms of experiential, physiological, and behavioural responses to meaningful stimuli (Mauss & Robinson, 2009), whereby the process of experiencing an emotion is referred to as *affect*. In this section, the tools available for measuring affect are considered. Following a general review of literature relating to emotions and affective processing (2.7.1), key methods relevant to addressing the research question are discussed. These methods include survey instruments (2.7.2), the startle reflex (2.7.3), and the International Affective Picture System (IAPS) (2.7.4).

2.7.1 Emotional Responses and Affective Processing

Emotional responses are incredibly varied and complex, however most theorists endorse an approach to emotion that features three components; "subjective feeling", "expressive behaviour", and "physiological arousal" (Scherer, 1993). Additionally, some add "motivational state", "action tendency" and/or "cognitive processing" (Scherer, 1993). While

these multiple emotional components are noted, simpler models to capture the motivational basis of emotion have evolved.

Many consider the motivational basis of emotion using a very simple, two-factor model featuring affective valence and arousal (Lang, 1995). This dimensional theory of emotion holds that all emotions can be located on a two-dimensional space, as a function of valence and arousal (Ravaja, Saari, Salminen, Laarni & Kallinen, 2005). In this two-dimensional model, valence represents a user's emotional reaction to a stimulus, reflecting the degree to which it is a pleasant or unpleasant experience (positive and negative valence, respectively). Arousal indicates the level of activation associated with the experience, from very excited and energized, to sleepy, calm, and/or disinterested (Ravaja et al., 2005). See Figure 3 below.

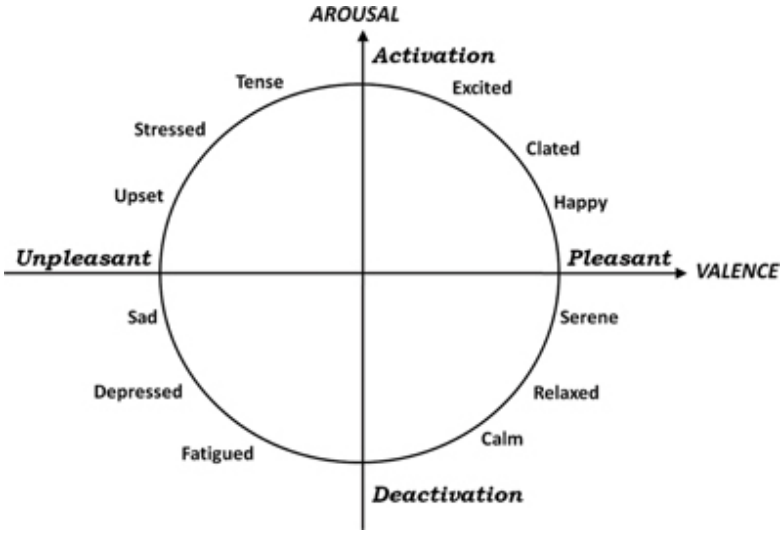


Figure 3 – Graphical representation of the valence/affect model (Valenza et al., 2011)

Affect, or affective processing, is bound in time to the experience of the world (stimuli) and the resulting emotional effect that this has (Barrett, Mesquita, & Ochsner, 2007). Affective processing is often a preferable construct than emotion, as the latter is more

prone to confusing and arbitrary definitions (Scherer, 2005). Additionally, affect is subconscious, and is a more reliable indicator of a person's core emotional state than self-reported emotion (Filion, Dawson & Schell, 1998). This aspect means that affect offers more opportunities for measurement than emotion, including collection of survey, observation and physiological data.

While the two factor model has received some criticisms for over-simplifying the emotional experience (Fontaine, Scherer, Roesch, & Ellsworth, 2007), the broad acceptance and usage of the valence/affect model make it a useful lens for understanding participant emotional responses to avatars. In this research, the focus is on the valence dimension due to the common perception that avatars may be unpleasant (see discussion of the Uncanny Valley in Section 2.6.) The measurement of emotional responses to stimuli can make use of both survey and physiological measurements, which are discussed below.

2.7.2 Survey Instruments (Godspeed Index)

Information obtained during the literature review on topics relevant to this study resulted in the location of material relating to a number methods to conduct an experiment to answer research questions similar to those posed in this research. Based on the literature, a measurement survey tool for an avatars perceived acceptance by viewers emerged. The study by Ho and MacDorman (2010), "Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices", provides details on a survey instrument highly relevant to this research. The original Godspeed Survey, developed by Bartneck (2008), is a tool that allows for the measurement of indices in the areas anthropomorphism, animacy, likability, perceived intelligence and perceived safety in the development of robots. Together, these dimensions provide a measure of the overall acceptance of a robot by users.

As this study is focusing on computer-generated avatars, the data gathering technique utilised is the alternate Godspeed indices developed by Ho and MacDorman (2010).

In their study, Ho and MacDorman (2010) discuss several issues with the original Godspeed instrument, such as the high level of correlation between anthropomorphism, animacy, likeability, and perceived intelligence. This high correlation indicates that some of these indices are measuring the same concept, and are thus are not appropriate for evaluating virtual anthropomorphic agents.

In light of these issues, the authors developed a modified version of the Godspeed survey that addressed the high correlations between dimensions, and also made it suitable for the study of avatars. It is this modified version of the Godspeed survey that is adopted in this research. In the modified version the original indices are replaced with 'humanness', 'eeriness' and 'attractiveness' indices. Ho and MacDorman (2010) describe the 'humanness' index as a measure of an avatars perceived or subjective humanness in order to check whether the 'objective' humanness has the intended effect. The 'objective' humanness relates to whether or not the avatars motion quality, skin texture or other property matches the expectation of the level of human-likeness that the avatar possess. 'Eeriness' is described as being distinct from negative warmth, which should be included in any set of indices related to works associated with the Uncanny Valley as it relates to the phenomenon. Finally, the 'attractiveness' index is explained as an important dimension in the Uncanny Valley explanation as it relates directly to the physical 'attractiveness' of an avatar. This 'attractiveness' is based on evolutionary cognitive and perceptual mechanisms for threat avoidance and mate selection.

These indices allow for measurement of a participants perception of an avatars 'humanness', 'eeriness' and 'attractiveness' making it a useful tool for designing and assessing avatars. For example, a study by Mitchell, Szerszen Sr, Lu, Schermerhorn, Scheutz,

and MacDorman (2011) has shown that it is an effective way to measure an avatars voice and facial realism, in order to determine the level of 'eeriness' experienced in their experiments. Using these indices this work created a design principle that the voice and visual elements of an avatar should match up in order to avoid the uncanny valley. In order to further understand participants' perceptions of avatars this survey was implemented in this research in conjunction with a startle response measurement in order to measure physiological reactions.

2.7.3 Physiological Measurement of Emotions (Startle Reflex)

Psychophysiological investigations suggest that some emotional states may be objectively measured with physiological measurements such as facial electromyography (EMG) (Dimberg, 1990), heart rate (Ravaja, Saari, Salminen, Laarni, & Kallinen, 2006) and galvanic skin response (Mundy-Castle & McKiever, 1953; Grimshaw, Lindley & Nacke, 2008). These three physiological measures are considered in this section.

Heart rate and galvanic skin response are the most commonly assessed indices of the autonomic nervous system (ANS) (Mauss & Robinson, 2009). Some studies of emotion use heart rate as a major physiological tool to gather data (Ravaja, Saari, Salminen, Laarni, & Kallinen, 2006), as it is established that tasks requiring cognitive effort or active coping elicit a heart rate (HR) acceleration that corresponds with emotional arousal, mediated by the sympathetic nervous system. However, given that the heart is dually related to both the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS), HR does not only measure emotional responses, but also attention, which may entail interpretive difficulties (Mauss & Robinson, 2009; Ravaja, Saari, Salminen, Laarni, & Kallinen, 2006).

Electrodermal activity (EDA), or skin conductance, is another physiological index of valence and arousal used frequently in the field of computer affect (Lang, 1993). As people experience emotions, their SNS is activated, resulting in increased sweat gland activity and

skin conductance. However, several studies have shown that increased EDA is highly correlated with emotional arousal of either valence, compared with lower-arousal pictures (Lang, 1993). These results suggest that EDA is measuring the response level of an emotional state rather than the specific emotional state itself (Mauss & Robinson, 2009).

According to Lang (1993), facial electromyography (EMG) is the primary physiological index of valence and thus makes it a relevant technique for this research. EMG activity increases with the contractions of various facial muscles that are closely linked with positive or negative emotions. It has been established that increased activity of the zygomaticus major (cheek) and orbicularis oculi (eye) facial muscles correspond to high valence conditions, whereas the corrugator supercilii (brow) muscle corresponds to low-valence conditions, during affective imagery (Witvliet & Vrana, 1995) (See Figure 4). The activity of these facial muscles is typically measured using startle reflex. The startle reflex is a defensive response to unexpected external stimuli, shown by contraction in skeletal and facial muscles. It has been extensively studied in rats and humans (Ramirez-Moreno & Sejnowski, 2012).

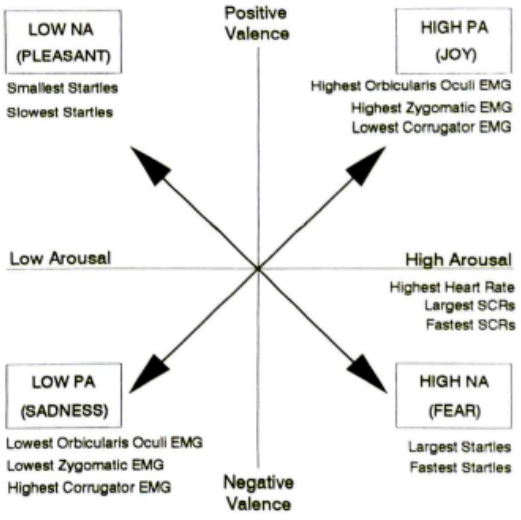


Figure 4 – Organisation of the startle reflex eyeblink and physiological indicators (Witvliet & Vrana, 1995, pg. 441)

The startle or blink reflex can be elicited using a variety of stimulus including acoustic, visual, electrical, magnetic and mechanical stimulation (Blumenthal et al., 2005). Acoustic stimulation uses a loud burst of white noise, delivered via headphones or speaker. Visual stimulation generates two reflexive blink responses; the photic blink in response to a sudden increase in illumination and the Cartesian blink as a result of a rapidly approaching object. Blink responses can also be triggered by stimulation of trigeminal cutaneous nerve fibres using electrical, magnetic or mechanical stimulation. In terms of measuring valence, acoustic stimulation is the principle method (Mauss & Robinson, 2009; Blumenthal et al., 2005).

Typically, participants who are being measured by an acoustic startle reflex probe are measured by placing an electrode situated under the left eye, below the lower eyelid in line with the pupil in a forward gaze (over the Orbicularis Oculi) (Blumenthal et al., 2005). A second electrode is also placed beside the first, and a third ground electrode is placed on an electrically inactive point such as the forehead or temple. Participants wear headphones, or listen to speakers, that deliver the startle pulse, which is a broadband noise delivered in intervals, usually containing frequencies in the 20 Hz to 20 kHz range at anywhere from 80 to 115 dB (Blumenthal et al., 2005). This application of the startle reflex probe will convey results that can be used to test a participant's emotional valence.

The magnitude of the eye blink response to the stimulus is interpreted as a user's aforementioned 'valence'. The more negatively perceived the stimulus is to the participant, the more significant the response (higher muscular activity), signalling a lower valence. In contrast, a more positively perceived stimulus will lead to a lesser response (lower muscular activity), signalling higher valence. Valence is often related to the participant's mood, with higher valence representing a more positive, 'happier' mood, whereas negative valence representing a lower mood (Frijda, 1986).

The startle reflex is recognised as having high variability among individuals (Blumenthal et al., 2005). In order to better understand an individual's startle reflex response to stimuli, baseline imagery with known response parameters are frequently used (Lang, Bradley & Cuthbert, 2007). The following section provides a discussion of one such set of imagery.

2.7.4 International Affective Picture System (IAPS)

The International Affective Picture System (IAPS) is commonly used to establish a baseline for each participants' individual affective processing in response to stimuli. The aim of this system is to aid in the experimental investigation of emotion and attention through a set of normative emotional stimuli. This is achieved by assembling a standardised set of emotionally-evocative imagery with a broad range of sematic categories that is internationally accessible. Lang, Bradley and Cuthbert (2007) argue that this collection provides higher level of experimental control in regards to the selection of the emotional stimuli. In addition this system can be utilised to compare results conducted within the same laboratory, or a different laboratory. Finally, IAPS can allow for exact replications across labs that are investigating all levels of psychological science.

The IAPS has been developed and distributed by the Centre for Emotion and Attention at the University of Florida. Specifically, the system has been constructed to provide indicative ratings of affect for the supplied sets of imagery. The images that make up the system have been presented to a global participant pool and rated on the three dimensions of *arousal*, *valence* and *dominance*. These ratings are indicated on the Self-Assessment Manikin (SAM), which is an affective rating system developed by Lang (1980). For the purpose of this study, images will be selected on the valence dimension alone. The primary focus of this research is the perception of pleasantness or unpleasantness associated with the

overall appearance and movement of developed avatars, making valence the appropriate measure.

2.8 Summary

For the purpose of this study, a formal systematic literature review was undertaken in order to gain a foundational understanding of the research area. The method suggested by Briner and Denyer (2012) was utilised in this review, with the results aided the formulatioin of this study's research questions. The systematic approach was followed by a narrative style review of literature appropriate methods and approaches. Together, the areas of discussion included avatar and human interaction, facial animation of avatars, the study of emotional expression in avatars, avatars levels of fidelity/realism, the Uncanny Valley, and tools for measuring emotional reactions.

From the systematic review, the following emerged as key points to inform the research:

* Interesting and important points relating to the impact of cultural diversity in the both the communication of emotional expression in avatars, and also in the ways that individuals from different cultural backgrounds may perceive avatars were identified in the literature. In light of this, some consideration to incorporating cultural diversity into this research is adopted, however given the complexity of the issues, it is not a focus of the research.
* Literature relating to facial animation of avatars indicated that there is no standardised method of generating any form of facial animation. Several of the techniques located in the literature were discussed, with automated markerless motion capture techniques emerging as both cost effective and efficient. As a result, an off the shelf markerless motion capture animation tool, Faceshift

(Version 2014.2.01, Faceshift AG, 2015), will be used in this research to create both still and animated avatars for experiments.

* Avatar fidelity and realism are poorly defined in the literature, and studies considering these concepts do not provide clarity on their impact on human perception in relation to emotional expressions. Areas such as discrete facial expression and eyes emerged as a common thread through the literature, focussing on how these areas can better aid avatar animation acceptance if done correctly. However, it was noted from a study undertaken by Noël, Dumoulin and Lindgaard (2009) that participants identified the emotional state of a face presented to them regardless of the level of realism of the avatar face. This is an important finding in the context of this research, as it implies that communication of emotions in virtual settings is not dependent on the human-likeness, or realism, of the avatar used to convey them.

* There was some evidence in the literature that identification of emotional expressions in avatars is sensitive to small changes in the presentation of avatar faces, suggesting that is important that there is a clear indication of what emotion an avatar is trying to express in order for effective and affective communication to occur for both the participant and avatar designer. This point is noted, and experiments in this research will use slightly exaggerated emotional expressions to avoid confusion.

* The importance of aligning visual and behavioral realism for increased avatar effectiveness was identified. Research identified that lower-realism avatars were not outperformed by those with higher levels of realism (Garau, Slater, Vinayagamoorthy, Brogni, Steed and Sasse, 2010). This suggests that the higher the realism of an avatar, the greater expectation there is for realistic

behaviors. Additionally, the expectation is that highly realistic facial movements seen in photorealistic avatars will impact on the overall user acceptance of the avatar, but to not add to the ability for the avatar to communicate emotions.

* An analysis of the literature located did not reveal any form of measurement for testing or evaluating levels of fidelity and realism in avatars. Many of the studies discussed did however give several areas that can contribute to levels of fidelity. These areas included the usage of wrinkles and facial colouring in relation to emotional expression as a form of increasing the level of avatar fidelity. As a result, the fidelity and realism of avatars developed for this research will be classified according to a scheme developed for this research. Details of these classifications are provided in the Methods chapter.

A narrative literature review approach was taken to explore literature relating to the Uncanny Valley, human emotional responses, and tools for measuring emotional reaction. Through this review, the following key points were identified:

* The motivational basis of emotion can be considered using a very simple, two-factor model featuring affective valence and arousal (Lang, 1995), where valence represents a user's emotional reaction to a stimulus, reflecting the degree to which it is a pleasant or unpleasant experience (positive and negative valence, respectively). Given the general perception that uncanniness is an unpleasant feature, valence emerges as they key emotional measure for this research.

* Originating from the study of uncanniness in robots, a number of valid survey instruments exist for measuring perceptions of uncanniness. As this study is

focusing on computer-generated avatars, the data gathering technique to be utilised is the alternate Godspeed indices developed by Ho and MacDorman (2010), which has been modified specifically for this purpose.

* Finally, physiological measures of emotional responses to visual stimuli were considered. Facial electromyography (EMG) of the human startle reflex was identified as the primary physiological index of valence, making it the relevant technique for this research.

In the following chapter of this thesis, the methods for the research undertaken will be discussed.

3 METHODS

3.1 Introduction

The literature discussed in the previous chapter identified three useful tools for measuring participants' emotional responses to visual stimuli. These tools included the International Affective Picture System (IAPS), human startle reflex, and an alternate Godspeed survey developed for avatars. This section provides details on how these tools were implemented in the experiments for this study in order to answer the research questions below. The discussion provided in this chapter also includes the methodology that was used to answer the identified research questions, including the specific research methods used to gather relevant information. Further discussion will include participant numbers and recruitment, experiment procedure and the collection and analysis of data. Finally, the limitations of the chosen research methodology will be analysed in the concluding section of this chapter.

3.2 Research Questions

In order to extend the existing knowledge surrounding avatars and their levels of fidelity and emotional expression in relation to the uncanny valley, this thesis aims to answer the following questions:

Research Question 1: ***How does avatar fidelity or realism influence the emotional experience of the interactions between humans and computer-generated avatars?***

Research Question 2: ***How do the emotional expressions of avatar facial features affect participants in relation to the levels of valence experienced?***

Based on a review of the literature relevant to these research questions, the following hypotheses were formulated:

H1: Emotional expression in higher fidelity avatars is less uncanny than that of lower fidelity avatars

H2: Emotional expression in abstract avatars is less uncanny than that of realistic avatars.

In the following section, the framework of research methodology used to answer these research questions is discussed.

3.3 Research Framework

The research framework consists of the overall approach to conducting the research, including the *lens* used to view the research problem, and the methods used to address it. Areas for discussion in this section include the methodological framework utilised for data collection from repeatable experiments, which alludes to the use of a positivism approach to undertaking this research. This discussion will also include how the experiments were undertaken in the context of approaches used in previous studies to answer similar questions. Another area of discussion will include the sample size used in this study's experiments. Finally, the use of physiological tools for measuring the emotional reaction of participants will be discussed.

The research framework of this study incorporates a methodological process that involves the objective analysis of perceptions of research participants to visual stimuli. Using a scientific method to answer the research questions posed is an accepted means of generating new knowledge relevant to the areas of the various disciplines and domains associated with this research. Mukerji and Albon (2015) argue that utilising a scientific

approach through quantitative methodologies will allow for measurement, quantifying and outlining the limits of a phenomenon. As this study focuses on quantitative data types through the process of experiments, a positivist model was used in order to observe and record phenomena through a systematic way in order to determine the underlying cause and effect of the phenomenon.

A positivist approach was adopted primarily due to the need to conduct reproducible experiments, which allowed an objective approach to be taken in order for different observers to be able to agree on the phenomena that is being observed (Mukerji & Albon, 2015). Positivism is, in its essence, about studying the world in an objective way, and viewing data as being independent of the observer. Information is collected under strict rules through direct observation, which is broken down to isolate elements and patterns that demonstrate cause-and-effect (Vaishnavi & Kuechler, 2007). As the experiments in this study can be replicated through having been conducted under a set structure, this study makes use of the positivist approach. A positivist approach values reliability, repeatability and generalisability (Vaishnavi & Kuechler, 2007). As the experiments used in this study fit within this criteria, a positivism approach was adopted for the research.

Information obtained during the systematic literature review on topics relevant to this study resulted in the location of material relating to various methods to conduct an experiment to answer research questions similar to those posed. Based on the literature, a measurement for an avatars perceived acceptance by viewers emerged. The study by Ho and MacDorman (2010), “Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices”, provides details on a survey instrument highly relevant to this research.

Similarly, a study by Weyers, Mühlberger and Pauli (2006) was used as the basis for determining duration of stimuli presentation to experiment participants. The length of time

that images will be displayed and how the avatars emotional expression would change within that set time frame came from this work. The guidelines for the measurement of the human startle reflex by Blumenthal et al. (2005) gave this study a clear process for the appropriate use of this approach. All of these elements were combined in this study to determine the sample size of participants in the research.

The population for this study will consist of both genders, with the age range expected to include persons 18 years of age and above. The sample size of participants for this study does not represent the entire population. Davis and Mosdell (2006) argue that due to strict time limitations placed upon projects it is difficult to gather a truly representative sample of the population. The purpose of individual research projects under the positivist paradigm is therefore to adopt a methodology that enables the findings to add to an existing body of knowledge that can be further explored in a consistent and valid manner. In the related areas of information technology, human-computer interaction, simulation and games production, producing results that are replicable is a desirable attribute of research undertaken in this field (Elson & Quandt, 2014; Wilson, Machay, Chi, Bernstein, Russell & Thimbleby, 2011; Richey 1998). This was the reasoning behind using sound methods and approaches that previous studies have used on similar research questions. This allows for a contribution of useful and accurate information that is applicable within these areas of study.

Consistency in the specific approach adopted is necessary to ensure replicability of results, and this is particularly true for the recording of the human startle, or eye-blink, response. In their 2005 article, “Committee report: Guidelines for human startle eyeblink electromyographic studies” Blumenthal, et al argue that any report using this approach should record all procedures in order for an effective evaluation of the methodology used to take place. This research complies with these guidelines by reporting all information related to the

experiment, such as how the stimuli were delivered and on what equipment it was delivered so that others may replicate this experiment.

While the fidelity of avatars, and the ability of avatars to produce accurate human emotional facial expressions are of interest in this research, it is the emotional state of participants in response to these avatars that provides the results to answer the research questions. Unfortunately, the goal of recognising emotional state is extremely challenging, as the concept of emotion is difficult to define as well as measure. This occurs as common moods, feelings, and attitudes vary significantly between individuals, both in how they are experienced and how they are expressed (Calvo & D'Mello, 2010).

As discussed in Chapter 2 of this thesis, an approach to measuring emotional states involves the physiological measurement of human startle eye-blink. The startle eye-blink is part of a non-voluntary response that typically occurs when an individual encounters a sudden and unexpected stimulus, such as a loud noise or increase in light (Blumenthal, Cuthbert, Filion, Hackley, Lipp, & Van Boxtel, 2005). In this study, participants are evaluated on the valence (pleasantness or unpleasantness) of still images, and short animation sequences, of a variety of avatars. These avatars vary in their level of fidelity and realism. The startle eye-blink amplitude, elicited using an acoustic white noise probe, allows for the measurement of participants' emotional reaction to these visual stimuli. In the following section, the justification of the methodology is presented.

3.4 Justification of the Methodology

In order to answer the research questions for this study, a mixture of quantitative and experimental methods are used to generate results from which to draw findings. This section will discuss the techniques used, specifically surveys and experiments. There will be an

analysis of the key studies relevant to this research, and an evaluation of the methods they adopted in terms of their utility for this research project.

The primary methods for collecting data to support the research questions posed by this study are experiments and surveys. Using a positivist view, this combination of experimental and quantitative data collection is used to gather objective results for statistical analysis. This analysis includes the use of a repeated measures ANOVA for testing data gathered using the Godspeed survey instrument, and is used in conjunction with descriptive statistics, T-Tests for comparison of means and correlations from the startle response data. Another set of T-Tests were used to examine if responses differed based on participant or avatar gender. Finally, a series of correlations was undertaken to determine if there were any correlations between the startle response and Godspeed data sets. There are several key studies that this research has examined to in order to develop the methodology for this research. These areas include how avatars affect humans, the various approaches other studies have used to evaluate fidelity and finally, methods for testing emotional expression in avatars.

The review of the literature found that computer generated avatars can affect humans. Based on studies by Philipp, Denney, Moncrieft, Sharma, and Vanman (2007) as well as a study from Wrzesien, Rodríguez, Rey, Alcañiz, Baños, and Vara (2015), this research seeks to use avatars to gain an emotional reaction from participants. The study by Philipp et at (2007) found that emotional behaviours occur in human within the presence of avatars. Wrzesien, Rodríguez, Rey, Alcañiz, Baños, and Vara's (2015) research shows that the emotional valence and arousal experiences are significantly more intense when observing an avatar that is physically similar to themselves. Based on these results this study will make use of computer generated avatars in order to gain a physiological reaction based on valence of participants.

One of the issues that will be considered in this study is user perception of the levels of avatar fidelity. A systematic review of literature revealed no standard way of generating or measuring fidelity. However some studies have reflected upon aspects that can aid the creation of higher-fidelity avatars. Steptoe and Steed, in their 2008 study "High-fidelity avatar eye representation" and Garau, Slater, Vinayagamoorthy, Brogni, Steed and Sasse in their 2003 study "The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment", reflect on how eye representation and gaze can greatly impact the level of avatar fidelity. Their results lead the authors to conclude that the higher the fidelity, the more expectation a viewer has on the avatar performing like a full-functioning human.

As there was no standard way of generating or measuring fidelity, this study made use of high-fidelity avatars already available in the public domain. This option was implemented due to time and resources constraints. There are many low-fidelity avatars also readily available that were utilised for the experiments. A classification scheme was developed for this research, which is detailed in Section 3.5.2.

The second issue that this study considers is the emotional response of participants in relation to the emotional expression displayed by avatars. The literature again revealed no standardised way of generating or testing emotional expression of avatars. However, a study by Weyers, Mühlberger and Pauli (2004) utilised electromyography (EMG) recording of the human startle reflex to record physiological reactions from participants in their study. This technique was adopted for this study, using recorded EMG signals of the startle reflex in response to an acoustic probe in order to determine a participant's emotional reaction to the visual stimuli presented to them. However, this is not the only technique that is being utilised for this study. In order to gain more insight from participants, this research makes use of an alternate Godspeed survey.

The original Godspeed survey is an instrument to measure users' perceptions of robots, based on the indices of anthropomorphism, animacy, likeability, perceived safety and perceived intelligence (Bartneck, Kulić, Croft, & Zoghbi, 2008). This study uses the alternative version developed by Ho and MacDorman (2010) that has been adapted for use with avatars. Specifically, the survey was adapted for the rating of avatars on the basis of 'attractiveness', 'eeriness' and 'humanness'. These alternate indices allow for better indications of participants perceptions of the visual stimuli presented to them. The survey instruments are discussed in more detail in Section 3.5.3. In the following section, the research methods used to provide answers to the research questions will be discussed.

3.5 Research Methods

This research was conducted with the approval of the University of Newcastle's Human Research Ethics Committee (H-2015-0163).The experiment procedure used in this research involves four distinct stages (Figure 5), namely Pre-Experiment, Experiment – Avatar Fidelity, Experiment – Emotional Expression, and Post-Experiment. Following a brief summary, the remaining sections provide details on the individual methods and considerations that relate to each of these stages.

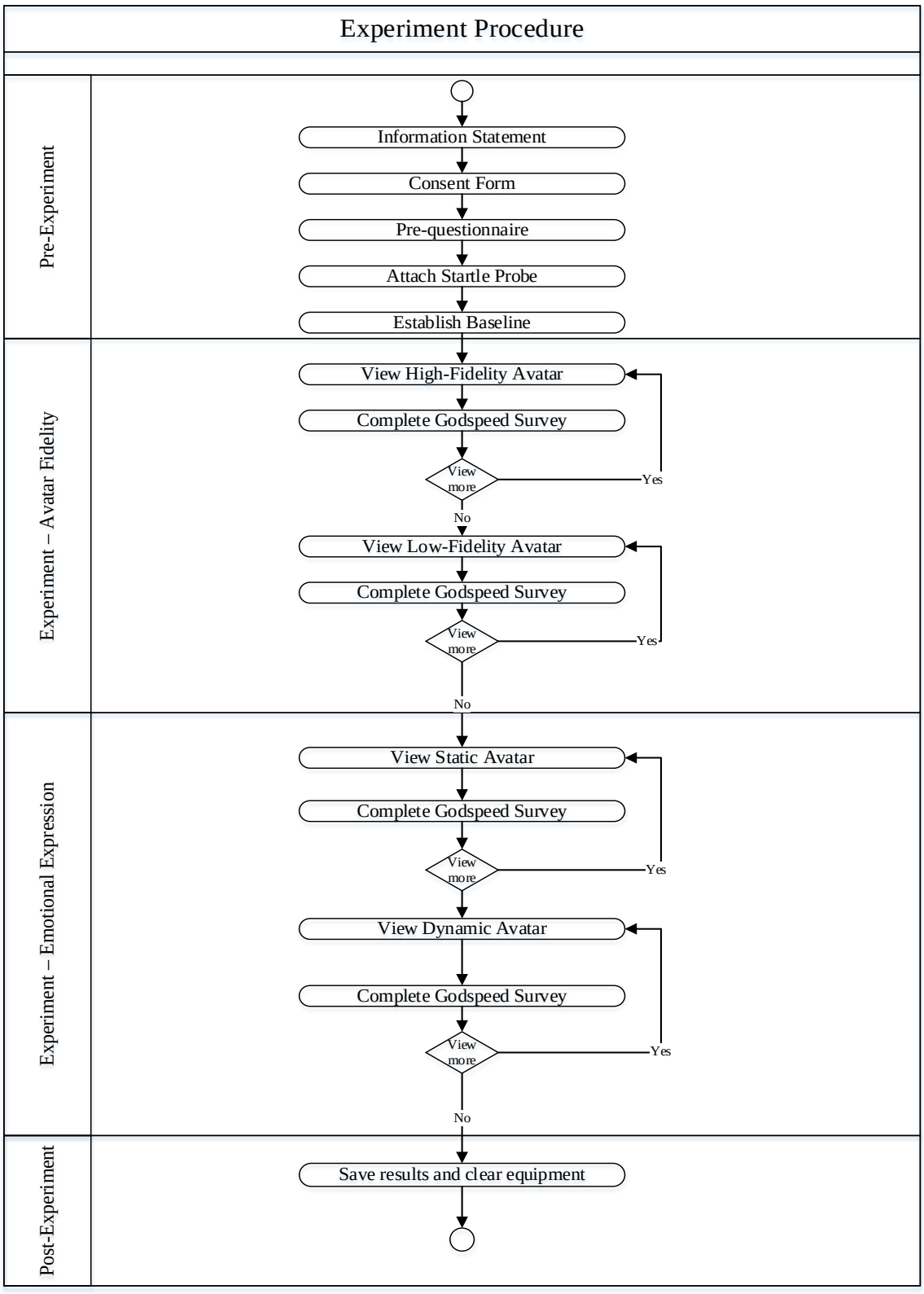


Figure 5 – Experiment Procedure

Prior to the commencement of the experiment, participants were required to complete a short demographic questionnaire (see Appendix A). In addition to standard questions relating to age, gender and educational background, this questionnaire also collected data on the familiarity participants have with the concept of the Uncanny Valley. In addition, data was gathered on what level of character animation experiences participants have.

In order to obtain effective and comparable measures of the human startle reflex, a series of baseline measurements were recorded prior to the commencement of the formal experiment. Once a baseline had been established, participants were asked to view a set of visual stimuli, in the form of animated avatars, related to the fidelity section of the experiment. After each viewing, participants were asked to complete a Godspeed survey on each image presented to them (see Section 3.5.3 and Appendix B). In the second half of the experiment, participants viewed a set of visual stimuli (avatars) for the emotional expression part of this testing. Once again, after the viewing of each image in this set, participants were asked to complete a Godspeed survey on each of the stimuli presented to them. The number of participants and the recruitment strategy adopted will be outlined in the next section.

3.5.1 Participant numbers and recruitment

This research aimed for a minimum sample size of 30. Thirty participants represents the minimum target for the study, with 50 being an upper bound within the allowable time. Based on Cresswell's (2007) suggestion of between 25-30 participants for a general study, a sample size of 30 provides a conservative estimate for this research. Additionally, Guest, Bunce and Johnson (2006) advise where the research requires comparison between distinct groups, the sample size will need to be larger, with each group being treated as a separate homogenous population.

The recruitment of participants was conducted primarily through posters placed on the University of Newcastle's Callaghan campus that outlined the project and provided contact information for the project researchers. These posters were displayed on notice boards around campus. Once contact between a participant and the researchers had been made, an appointment was arranged for the participant. Upon arrival at the appointment, the participant was provided with an information sheet that contained all the details of the project. Participants were asked to sign a consent form before any part of the experiment commenced. At any point in the recruitment process, participants were invited to ask questions of the researchers or withdraw from the study. The structure of the experiment will be discussed in the follow section.

3.5.2 Experiment

This study utilised a three part experiment design to answer the specified research questions. These parts consisted of: pre-experiment, the experiment relating to avatar fidelity and finally the experiment relating to emotional expression of avatar facial features. As discussed above, during the pre-experiment, participants were given an information statement that outlined the experiment. After reading the information presented, participants were asked to sign a consent form before any part of the experiment commences. Once consent is given, participants completed a short demographic survey prior to commencement (see Appendix A). Following this, the electrodes for EMG recording of startle responses were attached to the participant. Three electrodes were placed on the participants face following the guidelines for recording of the human startle reflex by Blumenthal et al. (2005). One electrode was positioned below the left eye in line with the pupil, resting below the lower eyelid in line with the pupil in a forward gaze over the orbiluaris oculi. The second electrode was positioned 1-2cms, depending on an individual participants' facial structure, parallel to the first electrode.

The final electrode was placed on the electrically inactive point behind the ear, which will act as an isolated ground, utilising Biogel Model 1090 Biopotential Contact Medium UFI Morro Bay, CA as a conductive gel. See Figure 6 for visual representation of these placements.

Figure 6 - EMG Electrode placement on participants

Note reprinted from: Measuring the effect of sound on the emotional and immersive experience of players in a video game: a case study in the horror genre pp. 36 By W. M. T. Coppins, 2015, Callaghan: University of Newcastle, 2015

The correct placement of the electrodes was confirmed by asking the participant to blink three times and observing the resultant EMG signal recording. Participants were then asked to wear K-55 stereo headphones attached to an Arcam FMJ AVR380 amplifier so that the startle noise would be administered without interference. Once correct electrode positioning was confirmed, participants were shown nine (9) images from the International Affective Picture System (IAPS) in order to establish a baseline of their emotional reactions to images that are categorised as pleasant, unpleasant or neutral. These images were displayed on a screen that was 1920x1080 pixels wide and 75” diagonally with participants

sitting approximately approximately 1.5 meters away from the screen seated facing the centre of the screen. A sample image from each of these categories is shown in Figure 7.

Pleasant	Neutral	Unpleasant
		

Figure 7 – Sample of images from selected categories from the IAPS

The International Affective Picture System (IAPS) has been developed by the Centre for the study of Emotion and Attention at the University of Florida. This system provides carries a set of emotionally evocative photographs across a semantic range. The images selected were categorised as pleasant, unpleasant and neutral, and participants were shown three (3) randomly selected images from each set. The chosen set of images was selected based on the measured valence rating, with a total of 10 pictures for each category to allow some variability in the actual images shown to participants. Similar to Bartholow, Bushman and Sestir (2006), images with a low rating of valence close 1 were selected for the unpleasant category. In contrast, images for the pleasant category had been chosen for their high rating of valence close to nine (9). While neutral category images rated around 4.5 were utilised. The average valence and arousal ratings of the images used in this study can be found in Table 5.

Table 5 – Valence and arousal ratings of images used in this study

Image Type	Valence Ratings	Arousal Ratings
Pleasant	8.14(1.33)	4.80(2.45)
Unpleasant	1.49(0.94)	6.77(2.18)
Neutral	4.59(1.41)	3.68(1.93)

Note IAPS, International Affective Picture System. Numbers in parentheses are standard deviations. The identification numbers (from the IAPS manual; Lang et al., 2001) for the neutral images were 1390, 2272,7920, 7011, 7031, 9210, 9401, 9150, 7110 and 7595. The pleasant images were 5830, 2340, 5210, 5760, 2070, 1440, 1460, 8190 1750 and 1710. The unpleasant images were 9940, 3053, 3080, 3015, 3266, 3005.2,9183, 6563, 3301 and 9635.1.

The images were presented to each participant using a randomised selection with no two images the same for a particular participant. The random selection of images selected for the participant to view was implemented to reduce bias that could potentially be associated with any particular selected image. In addition, a pseudorandom sequence was used to randomise the presentation order.

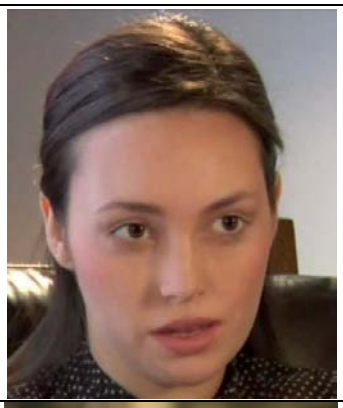
The ordering of the pictures was presented in a way that is similar to the approach used by Elsesser, Sartory and Tackenberg (2004). In their research, each participant was shown pictures in pseudorandom order and the same category was not presented twice in a row. The justification of the inclusion of pseudorandom ordering used in this experiment was used to minimise bias resulting from the previously viewed picture. During the presentation of these images, the startle probe/pulse was administered with timing of the noise varied in order to reduce familiarity.

In this study, in order to reduce the predictability, the startle probe pulses were delivered 3, 4 or 5 seconds after the picture onset. This method was used to reduce predictability with the startle pulses/probes in order to reduce any familiarisation that may occur which follows the approach used by Hess, Sabourin, and Kleck (2007). It is also similar to a method used by Bernat, Patrick, Benning and Tellegen (2006), who presented

images with startle pulses/probes occurring at either 3, 4 or 5 seconds after the picture was displayed.

The second part of this experiment tested participants' emotional reactions to levels of avatar fidelity. Participants viewed a high fidelity and a low-fidelity male avatar, as well as a set of corresponding female avatars. These avatars were displayed for a total of 5 seconds each, with the timing of the pulses to be similar to those used in establishing the baseline for the participants emotional reactions. Participants also completed an altered version of the Godspeed survey after viewing each avatar. The avatars used in this section of the experiments, as well as their fidelity category, can be found in Table 6 below.

Table 6 – Avatars used in the fidelity section of the experiment

Avatar	**Name**	**Category**
	**Emily** (*Source:* USC Institute for Creative Technologies Graphics Lab)	Fidelity = High
	**Ira** (*Source:* USC Institute for Creative Technologies Graphics Lab)	Fidelity = High

	**Liliwen** (*Source:* Created independently located on YouTube at https://www.youtube.com/watch?v=akdrHY9bBwk By nao4288 on 26th April, 2013)	Fidelity = Low
	**Leo** (*Source:* Created for research using FaceShift model)	Fidelity = Low

Finally, participants viewed a set of three avatars relating to the emotional expression of avatars. These avatars had a varying degree of realism, starting with the more abstract face of Curls, a mid-realism male and a mid-realism female. Curls and Macaw (mid-realism male) came packaged with the markerless motion-capture software FaceShift (Version 2014.2.01, Faceshift AG, 2015). While the mid-realism female (Jacqueline) was created in an online do-it-yourself software package from the 3D Avatar Store (https://3d-avatar-store.com/). Each of these faces was presented for 5 seconds with the pulses delivered similar to those discussed above. During presentation, the facial expressions changed from neutral to happy and neutral to sad, while participants were monitored by the startle probe apparatus. Between each avatar viewing, participants were asked to complete a Godspeed survey. The avatars used in this section of the experiments can be found in Table 7 below.

Table 7 – Avatars used for the emotional expression section of the experiments

Avatar	Name	Category
	**Jacqueline** (*Source:* Created for research using 3D Avatar Store, Maya, and FaceShift)	Fidelity = Low Realism = Mid
	**Curls** (*Source:* Created for research using FaceShift model)	Fidelity = Low Realism = Low
	**Macaw** (*Source:* Created for research using FaceShift model)	Fidelity = Mid Realism = Mid

In total, seven (7) human-like avatars were used for the experiments, four for the avatar fidelity (Table 6) experiment and three for the emotional expression experiment (Table 7). Categorisation of the avatar fidelity and realism levels was primarily based on the mechanism used to create the blend shapes or imagery for animation. In the case of the high fidelity avatars, these were created using a high resolution scanning process by a team of experts at the University of Southern California's (USC) Institute for Technologies Graphics Lab. In contrast, Leo, Curls and Macaw was created from inbuilt models in Faceshift

(Version 2014.2.01, Faceshift AG, 2015). Liliwen was sourced from an independent animator located within the public domain on YouTube, while Jacqueline was created indepentently. Both of these represent a 'do-it-yourself' style of character animation that can be achieved by an individual with the aid of off the shelf tools.

Only two categories of realism (i.e. mid and low) were defined. High realism was considered as an image or video of an actual human, thus the avatars were categorised based on their closeness to a real human. Avatars were considered to be of mid or medium level realism if they generated from images of real humans, whereas low realism avatars were created from artwork.

3.5.3 Surveys

For the purpose of this study, two surveys were administered to all participants. Firstly, participants were required to complete a demographic survey (Appendix A). This provided data on the demographic that has chosen to participate in the experiments for the research. Secondly, participants were required to complete the altered Godspeed survey for each avatar viewed in the experiment (Appendix B). The Godspeed survey was used in tandem with the recorded EMG data from a startle probe set up. This approach allowed for an evaluation of the data gathered independently, and was used to determine if the EMG data and the survey responses correlate. Robson (2002) states that surveys provide a straightforward method of obtaining data related to attitudes, motives and values of participants. Which lead to further analysis in order to establish what a participant thinks about an avatar in comparison to a physiological reaction to the presented stimuli.

Once the data had been collected from participants, the process of codification was undertaken. Data for participants was de-identified using an alphanumeric expression to identify records. For example, participants were identified as P0001 through to *P00n*,

depending on the number of participants overall. Gender was coded as either a 1 for female or 0 for male, which was then be used to generate a percentage of participant responders' gender. Age was recorded as a range between 18-25, 26-40, 41-55 and 55+ in order to calculate a histogram of the age range of all respondents.

For the survey questions related to educational background, avatar interaction, on what platform do participants interact with avatar, and for how many hours, responses were coded from 1-5, and a percentage was generated from each response. For the degree being studied, a percentage was used in order to gain a scope of the respondents' educational background. Finally, for the areas related to character animation and the concept of the uncanny valley, responses were coded from 1-3 and a percentage was be generated from the data obtained in order to assert if the participants are familiar with either of these areas. A copy of the demographic survey can be found in Appendix A.

In regards to the Godspeed survey, participants were again de-identified in the manner stated above. The three areas considered in participants' responses, that is 'humanness', 'eeriness' and 'attractiveness'. Responses for each of these were coded from 1-5. For each of the avatars, mean scores of the participant responses were calculated in order to give an average response to each of the criteria. A copy of the Godspeed survey can be found in Appendix B.

3.5.4 Collection and Storage of Data

The data collected from this study was obtained from the Godspeed survey and the startle probe EMG responses in form of quantitative data. All information collected was assembled and used with participant identities remaining anonymous. This data is kept in a locked storage facility at the School of Design, Communication and Information Technology. The period of retention is set at 5 years after which time all data will be destroyed. In

addition, data was stored in an area on a DCIT Sever; this secure server has firewalls and password encryption. The area for this data is only be available to the research team associated with this study, the only exception is for reasons required under law, and all data will remain in a de-identified manner. The DCIT Sever conforms to the University of Newcastle's protocols for data integrity, access and confidentiality. Dr. Karen Blackmore and Jacqueline Bailey had access to this data in order to complete the research as outlined in this study.

3.5.5 Analysis of Godspeed Survey Data

As discussed in Section 3.5.3, data obtained from the Godspeed survey was coded into values in order to generate statistics for comparison. This was achieved through calculating the mean scores for each avatar based on participants' responses to the categories of 'humanness', 'eeriness' and 'attractiveness'. These scores were then subjected to a repeated measures ANOVA test.

Urdan (2010), states that a Repeated Measures ANOVA is a useful statistical test to examine patterns of changes that occur over time for varying groups. The Repeated Measures ANOVA divides up the variance of the dependent variable, in this case the avatars, while the independent variables (fidelity, emotional expression) are tested for differences. The sources of differences come from variation within the mean scores generated and the within-subject variance analysis. Within the repeated measures ANOVA, differences between avatars were examined through a pairwise comparison using Bonferroni correction. This correction was used because the data was subjected to multiple testings that were performed simultaneously. Due to the small numbers of participants, Greenhouse-Geisser was used to test for within-subjects effects to determine statistical significance.

3.5.6 EMG Startle Reflex Data Analysis

EMG data recording of the startle reflex or eye-blink response to the acoustic probe was recorded through Labchart (Version 8.0.8, ADInstruments, 2015) with the processing and analysis taking place in Matlab (Version R2014b, The MathWorks Inc., 2014). IBM SPSS Statistics Standard (Version 23.0, IBM, 2011) was used for statistical analysis. Standard T-Tests were used to determine if there were significant differences in the mean scores from respondents to the avatars that being viewed. Part one of the experiment tested if there was a statistically significant difference between participants' perceptions of high-fidelity and low-fidelity avatars. The data obtained in part two established if there are statically significant differences in participants perceptions to avatar emotional expressions presented to them in the stimuli.

The main statistical comparison that was used in this study was a paired T-Test. This test compares two samples in which each value in one sample is seen to be a natural partner in the other sample. This was used to test the difference between the two values, variation of values that exist within the samples were taken into account during this test. The outcome of this test produced a number known as the "t-value". This T-Test was applied to the EMG data in order to determine if there were any statistically significant differences in participant responses. Another series of T-Tests was carried out on both the EMG and Godspeed data sets to determine if there were any gender differences in the responses obtained. These differences relate to the participants gender and the gender of the avatar, in order to identify differences in perception of the avatars based on gender.

The final series of tests were a series of correlations in order to determine if there were any correlations between the EMG and Godspeed results for participants. These correlations were performed in Excel (Version 15.0.4753.1003; Microsoft, 2013) and can be

found in Appendix E. The combination of these tests allowed for statistical analysis was used to determine whether or not the following hypothesises, developed from the review of literature, are accepted or rejected:

> *H1: Emotional expression in higher fidelity avatars is less uncanny than that of lower fidelity avatars*
>
> *H2: Emotional expression in abstract avatars is less uncanny than that of realistic avatars.*

In the final section of this report the limitations of this research methodology will be discussed.

3.6 Limitations of Research Methodology

The research methodology used for this study contains several limitations that potentially constrain the overall study. A major factor is the time restriction that is placed upon an Honours project. The number of participants likely to take part in this study was also a limitation. Finally, issues relating to experience of using tools for the measurement of physiological indicators, and those for avatar generation, pose limitations for this study.

3.6.1 Time restrictions

Due to the imposed time limit of the Honours program, the depth of analysis possible may be affected. In particular, obtaining ethics approval, recruiting participants and conducting experiments is a time consuming process. Time restrictions therefore impact on the number of experiments that can be undertaken within the timeframe of the Honours program. Another issue that has occurred is the absence of footage supplied from the University of Southern California in regards to high-fidelity avatars. While the imagery

required was sourced in the public domain, animations developed for the specific purpose of the research would have been optimal. Unfortunately, the time restrictions do not allow much room for delays, and a formal risk management process was in place to deal with risks associated with this research. A copy of the risk assessment of this project can be found in Appendix C.

3.6.2 Participant Numbers

As the sample size for the experiments in this study can be considered small, there is a level of limitation relating to the generalisability of the results that were generated. That is, the individuals within the sample may not be representative of the population. The sample size adopted was based on the minimum number of participants recommended by previous studies. In order to somewhat address this limitation, the methodological approach undertaken by this study allows for reproducibility allowing others to replicate and test the gathered results.

3.6.3 Limitations of choice relating to physiological indicators and avatar generation

The recording device used to measure the EMG response associated with the startle reflex that was used in this research is only one of many physiological measurement tools available. This equipment was deemed as appropriate for this study, however there is potential that different results will be obtained using different equipment. This is an accepted limitation for this project that has been mediated through the adoption of standard and recommended practices. Similar issues could occur with the stimuli (avatars) used in this study in regards to avatar generation.

Avatar generation for this study used the program Faceshift Studio (Version 2014.2.01, Faceshift AG, 2015), which is another limitation of this research. There are many programs

that are similar or arguably better, but this study has chosen to make use of this software. The program used only represents an instance of the available techniques for avatar generation and animation. However, the review of the literature did not reveal any studies that made use of more than one animation program for any given research. Also, variations are likely to be subtle and are outside of the research focus of emotional expressions.

3.7 Summary

Being able to generate avatars that are both realistic and emotionally expressive are two qualities that many fields would find beneficial. Techniques for measuring how the expressions affect viewers of various stimuli can assist in generating better avatars with less negative affects. The methodological approach used in this research was compiled so that the results produced were in a form that is most useful to these fields. The approach that was utilised is a positivist research paradigm that combines survey results and data collected from a physiological tool in order to test the research questions.

The results generated from the experiments and surveys address the research questions by allowing for statistical analysis of the data collected. Firstly, this data allows for further analysis in regards to the levels of avatar fidelity and its influence on the emotional experience that can be experienced between humans and computer generated avatars. Secondly, the degree of the pleasant or unpleasant nature of the emotional facial expressions of avatars as perceived by participants in relation to the Uncanny Valley can be further explored.

Exploration of these concepts occurred through the statistical comparison within the EMG and Godspeed data sets. The EMG dataset was also compared with the International Affective Picture System (IAPS) in order to best understand participants emotional experiences when interacting with the avatars presented to them. In addition the data was also

subjected to tests searching for gender differences before being tested for correlations. The results of these tests are presented in the following chapter.

4 RESULTS

4.1 Introduction

This section provides an overview of the experimental data collected for the project, and a brief summary of results from the data analysis. The experimental procedure detailed in Chapter 3 - Methods has provided a large amount of data for analysis. The collection of both physiological measurements and traditional survey data allow for independent analysis of these data, and also the analysis of possible correlations in the data. In this report the focus is on providing descriptive statistics of the experiment participants, and compare the mean responses of participants to the avatars viewed in the different conditions of the experiment.

Statistical analysis of data collected during the experiments was conducted using Excel (Version 15.0.4753.1003; Microsoft, 2013) and SPSS (Version 23; IBM, 2015). T-Tests were performed in Excel on the EMG data to determine if there were any statistically significant differences between the means data for the different experiment conditions. A series of Repeated Measures Analysis of Variance (ANOVA) were performed in SPSS to determine if there were any statistical significances within the data gathered from the Godspeed survey. The results of the analysis of data are presented in the following sections, including descriptive analysis of the participant information.

4.2 Participant Demographic Results

A total of 21 participants were recruited for the study. Various methods were used to recruit participants, including invitation posters on campus, email invitation to Information Technology (IT) undergraduate students, and word of mouth. The majority (52%) of participants were in the 18-25 year age bracket, followed by the 26-40 (29%) and the 41-60

(14%) age brackets (Figure 8). One participant opted to not provide age information. Forty three percent (43%) of participants identified as female, with 57% identifying as male.

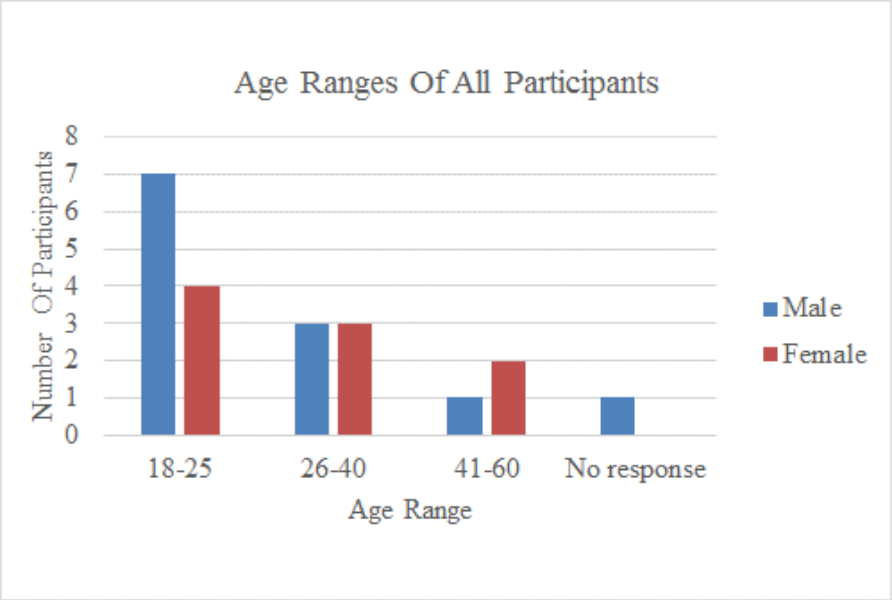


Figure 8 – Age ranges of all participants

Participants were from mixed educational backgrounds, with younger participants more likely to have either attained or be studying a Bachelor's degree. The number of hours per week spent engaging with avatars was variable. Interestingly, 57% of participants reported engaging with avatars of some form for 10 or less hours per week. A relatively large 19% of participants reported engaging with avatars more than 25 hours per week, indicating close familiarity with digital environments, as can be seen in Figure 9 below.

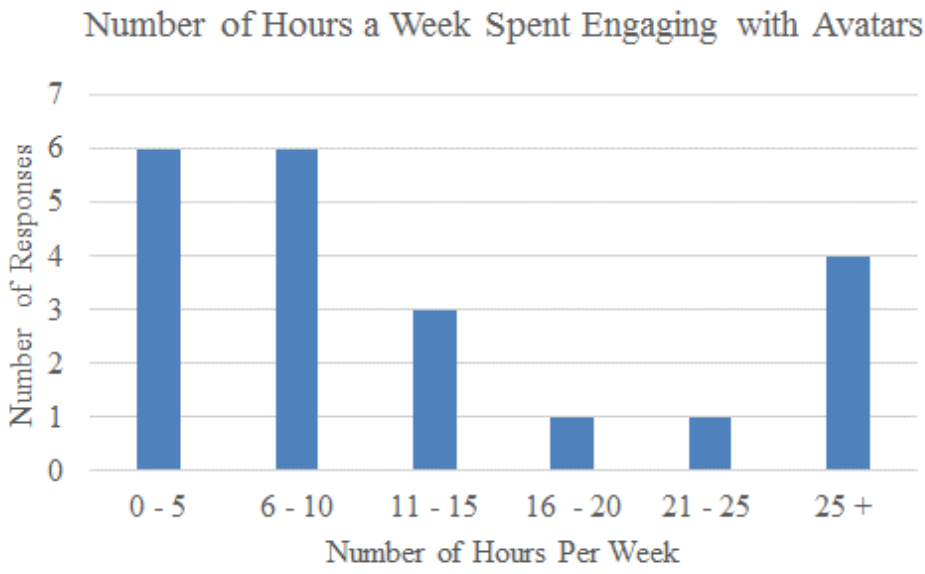


Figure 9 – Number of hours a week spent engaging with avatars for all participants

4.3 EMG Startle Results

The following sections present the results of participant EMG responses to a series of baseline still images, and also to the avatars under the varying experimental conditions.

4.3.1 EMG Startle Responses to the Baseline Images

A histogram of mean participant responses to unpleasant, pleasant and baseline imagery from the International Affective Picture System (Lang, Bradley, & Cuthbert, 2008) was generated to evaluate the consistency of participant responses (Figure 10). Some variability in mean responses is evident across all participants, however participant 13 (P0013) exhibits extreme results to pleasant imagery. As such, responses for P0013 are removed from the analysis of baseline results.

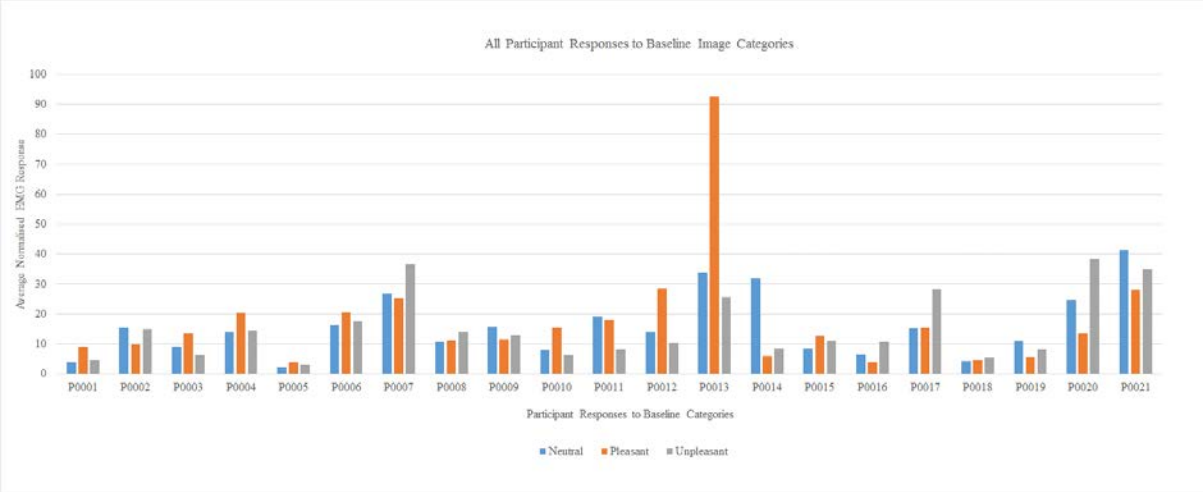


Figure 10 – All participant responses to the baseline image categories

The mean differences to neutral, pleasant and unpleasant baseline images are shown in Figure 11. The startle reflex was more strongly potentiated by pleasant images, than neutral or unpleasant images. The results for the pleasant versus unpleasant results are the reverse of those expected from the literature (Cuthbert, Bradley, & Lang, 1996; Lang et al., 1998; Vrana et al., 1988). Individual differences in startle responses are a known issue, and likely cause for these findings in this study due to the small sample size. Thus the EMG results should be interpreted with caution.

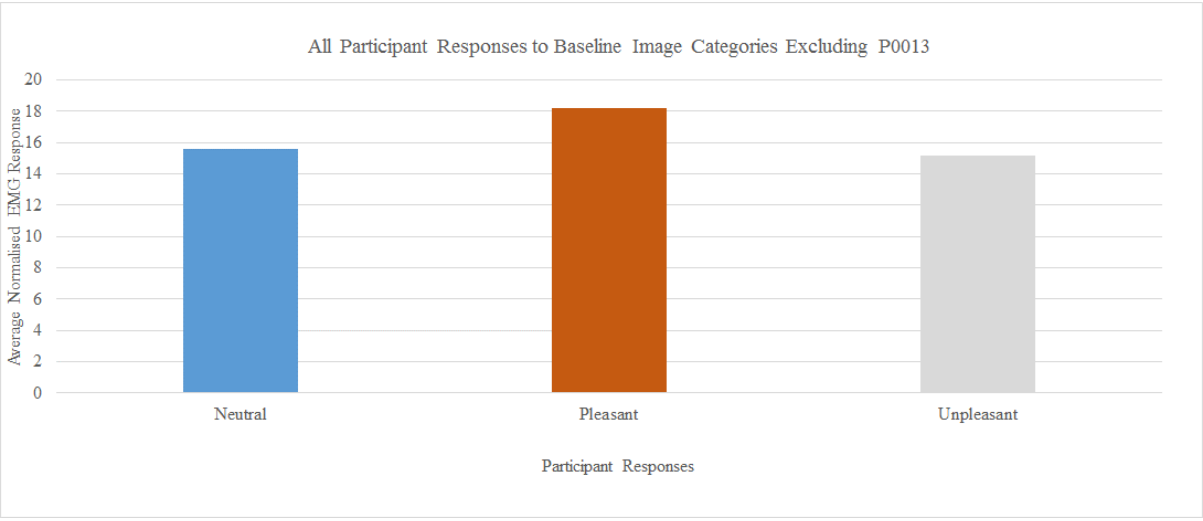


Figure 11 – All participant responses to the baseline image categories, excluding P0013

To explore this further, the EMG startle responses for male and female participants were compared (Figure 12). Interestingly, female participants startle reflex is uniformly more strongly potentiated that male participants, and male participant responses are higher for unpleasant than pleasant images, which matches findings from the literature. From these results, it is hypothesised that the gender variation, and in particular, the heightened EMG responses from female participants, is unduly influencing the overall averages.

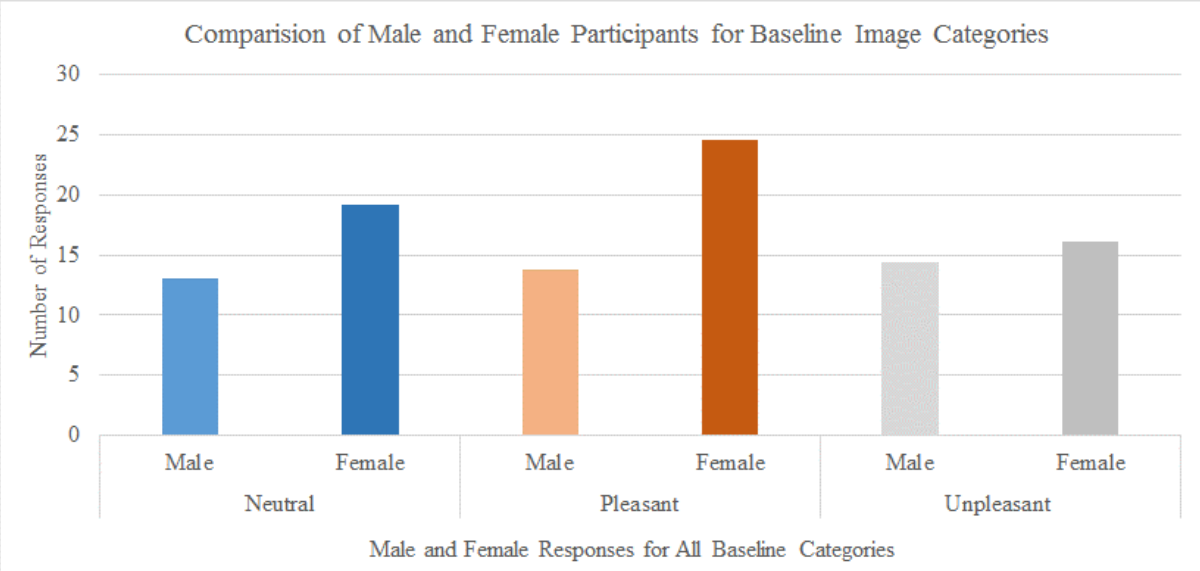


Figure 12 – A comparison of male and female participant response to all three categories of baseline imagery

Table 8 – T-Test results for EMG startle response

Test	Mean		Std. Dev		DF	p. value	t. value
Pleasant V Unpleasant	Pleasant	Unpleasant	Pleasant	Unpleasant	115	0.95	1.98
	13.99	14.17	11.84	19.34			
Pleasant V Neutral	Pleasant	Neutral	Pleasant	Neutral	112	0.75	198
	13.99	14.73	11.84	12.75			
Neutral V Unpleasant	Neutral	Unpleasant	Neutral	Unpleasant	123	0.84	1.97
	14.73	14.17	19.34	12.75			

The statistical differences in the mean responses for unpleasant, pleasant and neutral were examined using standard T-Tests and the results are shown in Table 8. Despite the apparent differences in the mean responses, the mean results were not statistically significant (i.e. p value > 0.05).

1.3.2 EMG Startle Responses to the Fidelity Avatars

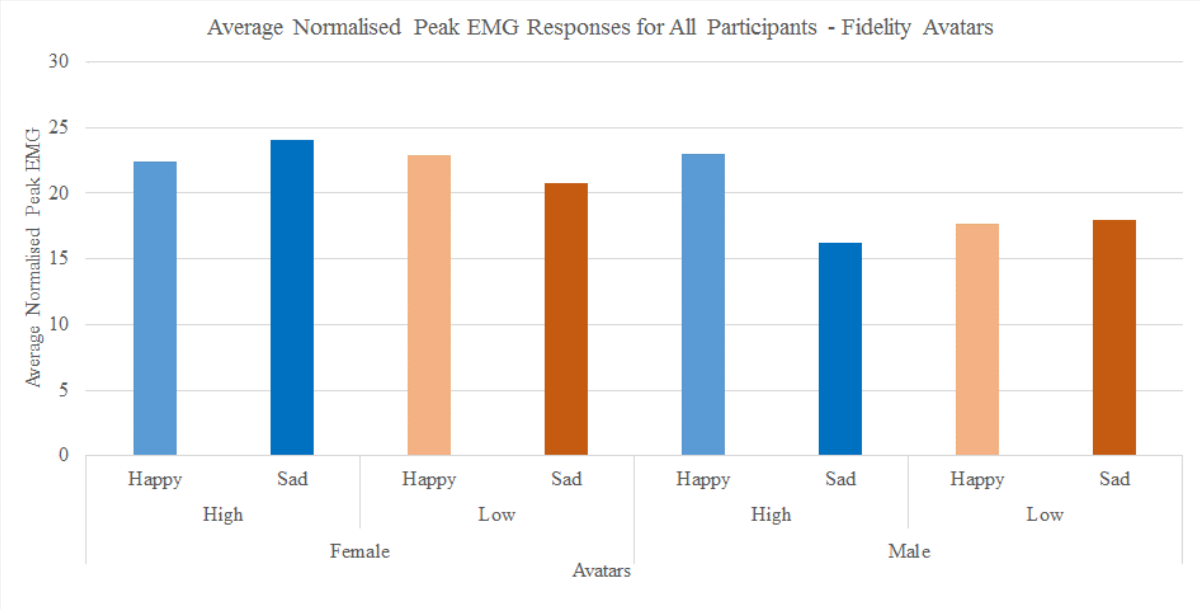


Figure 13 – Average normalised peak EMG for all participants viewing fidelity avatars

Figure 13 shows responses, as normalised peak EMG readings, for all participants when viewing the high and low fidelity avatars. The graph indicates that the responses were generally higher for both the high-fidelity and low fidelity female, which is indicative of higher arousal and more positive valence (intrinsic attractiveness) (VanOyen Witvliet & Vrana, 1995). These results are with the exception of the *Ira* avatar when displaying a happy expression, which appears comparable to results for the female avatar.

Table 9 - T-Test results for EMG startle responses for Fidelity Avatars

Test	Mean		Std. Dev		DF	p. value	t. value
High fidelity Male V High fidelity Female	**Female**	**Male**	**Female**	**Male**	77	0.38	1.99
	25.18	19.70	34.58	19.22			
Low fidelity Male V Low fidelity Female	**Female**	**Male**	**Female**	**Male**	74	0.40	1.99
	23.30	18.49	28.82	20.65			
High fidelity Female V Low fidelity Female	**High**	**Low**	**High**	**Low**	74	0.79	1.99
	25.18	23.30	34.58	28.82			
High fidelity Female Happy V High fidelity Female Sad	**Happy**	**Sad**	**Happy**	**Sad**	36	0.58	2.02
	22.39	28.64	26.77	42.97			
Low fidelity Female Happy V Low fidelity Female Sad	**Happy**	**Sad**	**Happy**	**Sad**	36	0.80	2.02
	24.50	22.11	25.98	32.08			
High fidelity Male V Low fidelity Male	**High**	**Low**	**High**	**Low**	77	0.78	1.99
	19.70	18.49	19.22	20.65			
High fidelity Male Happy V High fidelity Male Sad	**Happy**	**Sad**	**Happy**	**Sad**	39	0.23	2.02
	23.40	16.18	22.84	14.72			
Low fidelity Male Happy V Low fidelity Male Sad	**Happy**	**Sad**	**Happy**	**Sad**	36	0.96	2.02
	18.36	18.62	17.52	23.87			
High fidelity female V High fidelity male	**Female**	**Male**	**Female**	**Male**	77	0.38	1.99
	25.18	19.70	34.58	19.22			
Low fidelity female V Low fidelity male	**Female**	**Male**	**Female**	**Male**	74	0.40	1.99
	23.30	18.49	28.82	20.65			
High fidelity female sad V high fidelity male sad	**Female**	**Male**	**Female**	**Male**	36	0.22	2.02
	28.64	16.18	42.97	14.72			
High fidelity female happy V high fidelity male happy	**Female**	**Male**	**Female**	**Male**	39	0.89	2.02
	22.39	23.40	26.77	22.84			
Low fidelity female sad V low fidelity	**Female**	**Male**	**Female**	**Male**	36	0.70	2.02
	22.11	18.62	32.08	23.87			

male sad							
Low fidelity female happy V low fidelity male happy	**Female**	**Male**	**Female**	**Male**	36	0.39	2.02
	24.50	18.36	25.98	17.52			

The statistical differences in the mean high and low fidelity avatars, as well as differences between expression displayed (happy or sad) were examined using standard T-Tests and the results are shown in Table 9. The differences in the mean results were not statistically significant at the 5% level (i.e. p value > 0.05).

4.3.2 EMG Startle Responses to the Emotional Expression Avatars

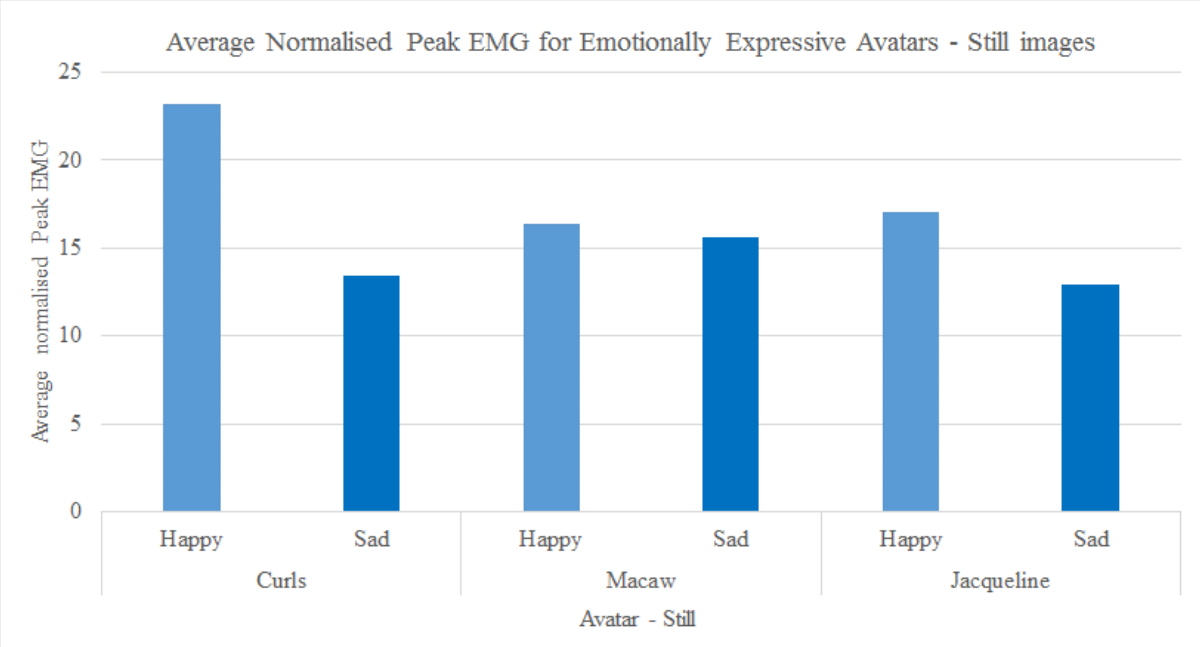


Figure 14 - Average Normalised Peak EMG for Emotionally Expressive avatars – still images

Figure 14 shows the average normalised peak EMG readings for the still images of avatars presented to all participants. The graph indicates that the still image of *Curls* displaying a happy expression gained the highest response, or the highest valence, which is associated with unpleasant imagery (Witvliet & Vrana, 1995). Thus, this can be interpreted as participants perceived this avatar and expression as more unpleasant that the others viewed.

Interestingly, Curls displaying a sad expression attained the second lowest response, suggesting that the facial expression, in combination with the actual avatar, has a substantial impact on participant perceptions, and this is most evident in avatars with low levels of realism. Overall, happy facial expressions appear to record a higher response, or stimulate a higher valence response (unpleasant) than sad expressions.

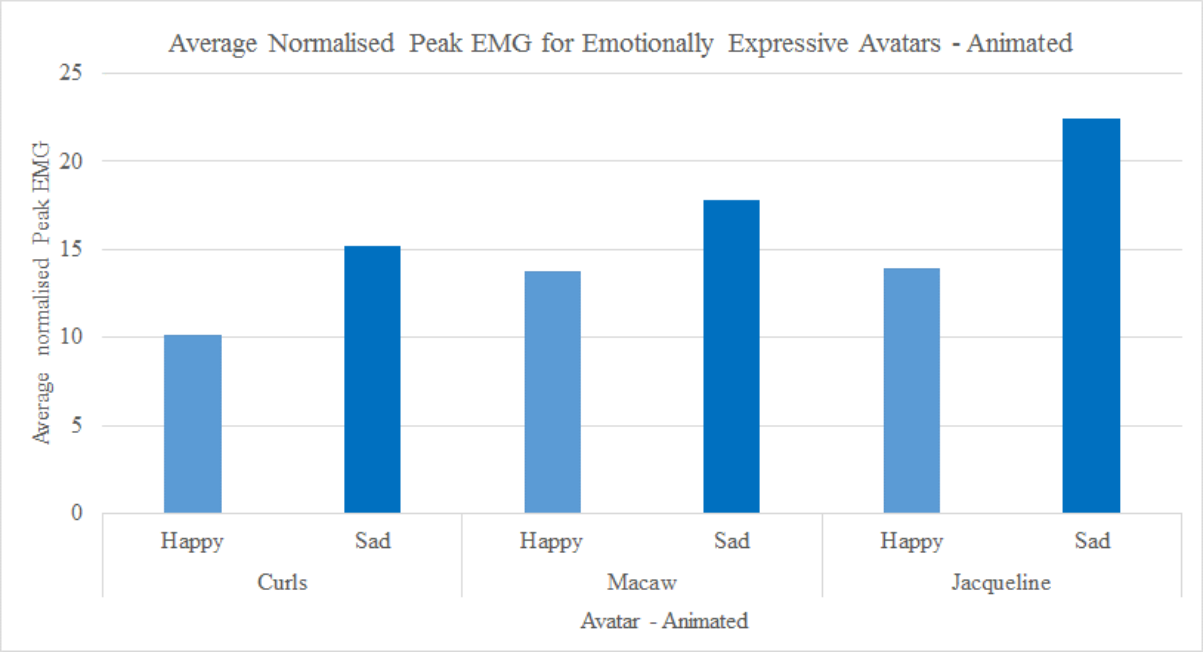


Figure 15 - Average Normalised Peak EMG for Emotionally Expressive avatars – Animated imagery

Figure 15 shows the average normalised peak EMG for the emotionally expressive avatars that were animated and presented to all participants. In contrast to the still images graph above, Curls happy has the lowest response in comparison to the other avatars. This low potentiation of the startle reflex suggests that avatars with low realism may not be as effective at communicating emotions as still images. In direct contrast to the still image of Jacqueline displaying a sad expression having the lowest response rate, an animated version displaying the same emotion has the highest response rate in comparison to the other avatars.

Again, these responses represent a participant's response on the valence dimension, and thus higher responses are indicative of a stronger, unpleasant emotional response.

Table 10 - T-Test results for EMG startle responses for emotional expression avatars

Test	Mean		Std. Dev		DF	p. value	t. value
Curls Still V Curls Dynamic	**Still**	**Dynamic**	**Still**	**Dynamic**	40	0.15	1.99
	18.31	12.69	21.35	12.99			
Curls Still Happy V Curls Still Sad	**Happy**	**Sad**	**Happy**	**Sad**	38	0.15	2.02
	23.16	13.45	25.60	15.18			
Curls Dynamic Happy V Curls Dynamic Sad	**Happy**	**Sad**	**Happy**	**Sad**	40	0.20	2.02
	10.14	15.23	8.56	16.09			
Macaw Still V Macaw Dynamic	**Still**	**Dynamic**	**Still**	**Dynamic**	82	0.96	1.98
	15.97	15.81	17.11	13.52			
Macaw Still Happy V Macaw Still Sad	**Happy**	**Sad**	**Happy**	**Sad**	40	0.88	2.02
	16.35	15.60	20.06	14.04			
Macaw Dynamic Happy V Macaw Dynamic Sad	**Happy**	**Sad**	**Happy**	**Sad**	40	0.33	2.02
	13.77	17.84	12.38	14.59			
Jack Still V Jack Dynamic	**Still**	**Dynamic**	**Still**	**Dynamic**	82	0.45	1.98
	14.96	18.15	16.37	22.02			
Jack Still Happy V Jack Still Sad	**Happy**	**Sad**	**Happy**	**Sad**	40	0.41	2.02
	17.03	12.89	20.32	11.29			
Jack Dynamic Happy V Jack Dynamic Sad	**Happy**	**Sad**	**Happy**	**Sad**	40	0.21	2.02
	13.90	22.41	11.16	28.83			
Animated Imagery T-Tests							
Curls Happy V Macaw Happy	**Curls Happy**	**Macaw Happy**	**Curls Happy**	**Macaw Happy**	40	0.22	2.02
	10.14	13.77	8.56	12.38			
Curls Happy V Jack Happy	**Curls Happy**	**Jack Happy**	**Curls Happy**	**Jack Happy**	40	0.27	2.02
	10.14	13.60	8.56	11.16			
Curls Sad V	**Curls**	**Macaw**	**Curls**	**Macaw**	40	0.58	2.02

Macaw Sad	**Sad**	**Sad**	**Sad**	**Sad**			
	15.23	17.84	16.09	14.59			
Curls Sad V Jack Sad	**Curls Sad**	**Jack Sad**	**Curls Sad**	**Jack Sad**	40	0.32	2.02
	15.23	22.41	16.09	28.83			
Macaw Happy V Jack Happy	**Macaw Happy**	**Jack Happy**	**Macaw Happy**	**Jack Happy**	40	0.97	2.02
	13.77	13.90	12.38	11.16			
Macaw Sad V Jack Sad	**Macaw Sad**	**Jack Sad**	**Macaw Sad**	**Jack Sad**	40	0.52	2.02
	17.84	22.41	14.59	28.83			
Still Imagery T-Tests							
Curls Happy V Macaw Happy	**Curls Happy**	**Macaw Happy**	**Curls Happy**	**Macaw Happy**	39	0.34	2.02
	23.16	16.35	25.60	20.06			
Curls Happy V Jack Happy	**Curls Happy**	**Jack Happy**	**Curls Happy**	**Jack Happy**	39	0.39	2.02
	23.16	17.03	25.60	20.32			
Curls Sad V Macaw Sad	**Curls Sad**	**Macaw Sad**	**Curls Sad**	**Macaw Sad**	39	0.64	2.02
	13.45	15.60	15.18	14.04			
Curls Sad V Jack Sad	**Curls Sad**	**Jack Sad**	**Curls Sad**	**Jack Sad**	39	0.89	2.02
	13.45	12.89	15.18	11.29			
Macaw Happy V Jack Happy	**Macaw Happy**	**Jack Happy**	**Macaw Happy**	**Jack Happy**	40	0.91	2.02
	16.35	17.03	20.06	20.32			
Macaw Sad V Jack Sad	**Macaw Sad**	**Jack Sad**	**Macaw Sad**	**Jack Sad**	40	0.49	2.02
	15.60	12.89	14.04	11.29			

The statistical differences in the mean differences between expressions displayed (happy or sad) and between differences in type (still or animated) were examined using standard T-Tests and the results are shown in Table 10. Again, despite clear differences in the mean results, these were not found to be statistically significant (i.e. p value > 0.05). Possible reasons for non-significant results are provided in the latter discussion section.

4.4 Godspeed Results

Each participant was requested to complete a Godspeed survey for every avatar presented to him or her. In total, participants completed 20 surveys each, with each survey consisting of semantic differential scale measures. The scales in the surveys were divided into three indices to measure the participants' perception of each avatar's 'humanness', 'eeriness' and 'attractiveness'.

In order to test for statistical significance, the data collected from the Godspeed survey was analysed by a Repeated Measures Analysis of Variance (ANOVA) in the SPSS (Version 23; IBM, 2015).

4.4.1 Repeated Measures ANOVA for Fidelity avatars

4.4.1.1 Humanness

A repeated measures ANOVA determined that the means for participant's perceptions of 'humanness' of high and low fidelity, male and female avatars (Ira, Emily, Leo and Liliwen), were significantly different. ($F_{(2.40,47.90)}$=85.57 p<.001) and similarly for their expressions (happy/sad) ($F_{(1.00, 20.00)}$ =7.00, p=.016), but not for the interaction of avatar and expression. See Table 11.

Post hoc analysis revealed that the differences were between Ira and Emily (p<.001), Ira and Leo (p<.001), Ira and Liliwen (p=.003), Emily and Leo (p<.001), Emily and Liliwen (p<.001), but not between the Leo and Liliwen. Therefore, Emily was perceived as the most human and both Ira and Emily were perceived as more human than both Leo and Liliwen. Also, avatars with sad expressions were perceived as statistically significantly more human than avatars with happy expressions (p=.016).

Table 11 - Godspeed means and standard errors for perception of humanness of fidelity avatars

Type	Avatar	Happy Mean(SE)	Sad Mean(SE)
High-Fidelity Male	Ira	2.7(0.24)	3.0(0.24)
High-Fidelity Female	Emily	4.4(0.15)	4.5 (0.15)
Low-Fidelity Male	Leo	1.6 (0.14)	1.7 (0.16)
Low-Fidelity Female	Liliwen	1.8 (0.16)	1.9 (0.17)
Overall		2.6 (.11)	2.8(0.13)

4.4.1.2 Eeriness

There were no statistically significant differences between participants' perceptions of 'eeriness' for any of the high or low fidelity avatars with either happy or sad expressions. See Table 12 below.

Table 12 - Godspeed means and standard errors for perception of eeriness of fidelity avatars

Type	Avatar	Happy Mean(SE)	Sad Mean(SE)
High-Fidelity Male	Ira	2.6(0.11)	2.7(0.12)
High-Fidelity Female	Emily	2.3(0.14)	2.4(0.14)
Low-Fidelity Male	Leo	2.5(0.17)	2.7(0.17)
Low-Fidelity Female	Liliwen	2.4(0.18)	2.4(0.16)
Overall		2.5(0.10)	2.5(0.10)

4.4.1.3 Attractiveness

The mean participant's perceptions of 'attractiveness' in high and low fidelity, male and female avatars were significantly different. ($F_{(2.50,\ 50.12)}$=26.38,p<.001), but not their expression, and not the interaction between avatar and expression.

Post hoc evaluation revealed there were differences of perceived 'attractiveness' between Ira and Emily (p<.001), Emily and Leo (p<.001) Emily and Liliwen (p<.001), and Liliwen and Leo (p=.038). There were no other differences. See Table 13. Therefore, Emily is perceived as being the more attractive, despite her expression, than all the other avatars,

and Liliwen is perceived as more attractive than Leo (i.e. both females were perceived as attractive).

Table 13 - Godspeed means and standard errors for perception of attractiveness of fidelity avatars

Type	Avatar	Happy Mean(SE)	Sad Mean(SE)
High-Fidelity Male	Ira	2.6(0.18)	2.7(0.18)
High-Fidelity Female	Emily	4.1(0.15)	4.1(0.19)
Low-Fidelity Male	Leo	2.3(0.21)	2.3(0.19)
Low-Fidelity Female	Liliwen	3.0(0.17)	3.0(0.21)
Overall		3.0(0.12)	3.0(0.13)

4.4.2 Repeated Measures ANOVA of Still and Animated Avatars

4.4.2.1 Humanness

The repeated measures ANOVA determined that the means for participant's perceptions of 'humanness' of the 3 low and mid realism male and females avatars (Curls, Macaw, Jacqueline), were significantly different, as was their type of imagery ($F_{(1.00,18.00)}$=19.02,p<.001) (still/animated). The interaction of avatar and expression (happy/sad) ($F_{(1.87,\ 33.69)}$=5.41,p=.010) was significant, but not the interactions of avatar and type of imagery used or avatar and type of imagery and expression.

Post hoc evaluation revealed that the difference was between Macaw and Jacqueline (p=.004) and between the participants' perception of the 'humanness' of the still and animated avatars (p<.001). There were no other differences. See Table 14. Macaw with a happy expression was perceived as the most human.

Table 14 - Godspeed means and standard errors for perception of humanness of still and animated avatars

		Still		Animated	
Avatar	Type	Happy	Sad	Happy	Sad
Curls	Low Realism	1.5 (0.13)	1.7(0.17)	1.9 (0.17)	1.9 (0.18)
Macaw	Mid Realism	2.0(0.21)	1.9(0.21)	2.2(0.21)	2.2(0.22)
Jacqueline	Mid Realism	1.4(0.09)	1.3(0.09)	1.8(0.15)	1.6(0.11)
Overall		1.6 (0.12)		1.9 (0.15)	

4.4.2.2 Eeriness

There were no statistically significant differences between the participants perception of 'eeriness' in relation to the avatars presented to them. See Table 15.

Table 15 - Godspeed means and standard errors for perception of eeriness of still and animated avatars

		Still		Animated	
Avatar	Type	Happy	Sad	Happy	Sad
Curls	Low Realism	2.6(0.21)	2.7(0.17)	2.9(0.20)	2.6(0.17)
Macaw	Mid Realism	2.6(0.22)	2.7(0.25)	3.0(0.19)	2.6(0.22)
Jacqueline	Mid Realism	2.3(0.20)	2.6(0.18)	2.5(0.18)	2.5(0.18)
Overall		2.6(0.14)		2.6(0.14)	

4.4.2.3 Attractiveness

The repeated measures ANOVA determined that the means for participant's perceptions of 'attractiveness' of the 3 low and mid realism male and females avatars (Curls, Macaw, Jacqueline), were significantly different, as was their type of imagery ($F_{(1.00, 19.00)}$=5.39,p=.032) (still/animated). The interaction of avatar and expression (happy/sad) ($F_{(1.73,32.88)}$=4.83,p=.018) was significant, but not the interactions of avatar and type of imagery used or avatar and type of imagery and expression.

Post hoc evaluation revealed that the difference was between Curls and Jacqueline (*p*=.003), Macaw and Jacqueline (*p*=.022) and between the participants' perception of the 'attractiveness' of the still and animated avatars (*p*=.032). There were no other differences. See Table 16.

Table 16 - Godspeed means and standard errors for perception of attractiveness of still and animated avatars

		Still		Animated	
Avatar	Type	Happy	Sad	Happy	Sad
Curls	Low Realism	2.6(0.25)	2.90.23)	2.8(0.28)	2.9(0.24)
Macaw	Mid Realism	2.4(0.18)	2.1(0.18)	2.4(0.18)	2.5(0.20)
Jacqueline	Mid Realism	1.9(0.19)	1.8(0.17)	2.1(0.18)	2.1(0.20)
Overall		2.3(0.16)		2.3(0.15)	

4.5 Other Findings – Gender Differences

4.5.1 EMG Data: *Baseline*

As stated above, and presented in Figure 11, the differences noted in the female and male EMG responses for the baseline images were interesting, and further analysis was conducted. Gender differences were considered for both the participants' gender and the gender of the avatar being viewed. The results indicate that there are no statistically significant differences between participant genders, despite substantial differences between male and female mean responses to pleasant and neutral images. See Table 17 below.

Table 17 – Means and standard deviations for female and male participants' baseline responses

Test	**Mean**		**Std. Dev**		**DF**	**p. value**	**t. value**
Females V Males - Unpleasant	**Female**	**Male**	**Female**	**Male**	65	.73	1.99
	15.55	13.94	20.15	18.69			
Females V Males - Pleasant	**Female**	**Male**	**Female**	**Male**	54	.07	2.00
	24.55	13.76	31.79	11.80			
Females V Males - Neutral	**Female**	**Male**	**Female**	**Male**	62	.07	1.99
	19.15	13.04	15.48	11.82			

4.5.2 EMG Responses to Fidelity Avatars – Gender Differences

T-Tests were performed on the EMG data gathered to determine whether or not there were any differences in results between the genders in regards to the level of fidelity of the avatars, and the gender of the avatar.

The results show that there is a statistically significant difference between female and male participants EMG results to the low-fidelity avatars ($p. < .001$), with the mean scores for females (M=29.09, SD=30.67) scoring higher than their male counterparts (M=15.25, SD=18.60). Additionally, average responses of participants to avatars of different genders, that is, Ira ($p. < .001$) and Liliwen ($p. = .004$), were also different.

All other tests for high and low fidelity avatars and genders were found to not be statistically significant. These test results, as well as the means and standard deviations, can be found in Table 18 below.

Table 18 - Results from T-Tests based on participant gender for Fidelity avatars

Test	Mean		Std. Dev		DF	p. value	t. value
Participants							
Female participants V Male participants – High Fidelity	**Female**	**Male**	**Female**	**Male**	77	.16	1.99
	21.51	18.63	25.19	28.98			
Female participants V Males participants – Low Fidelity	**Female**	**Male**	**Female**	**Male**	74	.01 *	1.99
	29.09	15.25	30.67	18.60			
Female participants V Male participants – High Fidelity Male Avatar	**Female**	**Male**	**Female**	**Male**	39	.00 *	2.02
	29.18	12.99	22.28	13.56			
Female participants V Male participants – High Fidelity	**Female**	**Male**	**Female**	**Male**	36	.93	2.02
	25.74	24.78	28.60	39.02			

Female Avatar							
Female participants V Male participants – Low Fidelity Female Avatar	**Female**	**Male**	**Female**	**Male**	36	.04 *	2.02
	34.84	15.78	37.07	19.29			
Female participants V Male participants – Low Fidelity Male Avatar	**Female**	**Male**	**Female**	**Male**	36	.18	2.02
	23.70	14.70	23.11	18.28			

Note: * indicates statistical significance at 5% level, i.e. (*p*. < .050)

Emotionally Expressive Avatars – Gender Differences

T-Tests were performed on the EMG data gathered to determine whether or not there were any differences in results between the genders in regards to the type of imagery (still or animated) of the avatars and the gender of the avatar.

The results indicate that there is a statistically significant difference between female and male participants when viewing an animated version of the Jacqueline avatar (*p*. = .004) All other results were not statistically significant, the means and standard deviations and test results can be found below in Table 19.

Table 19 - Results from T-Tests based on participant gender for Emotionally Expressive avatars

Test	**Mean**		**Std. Dev**		**DF**	**p. value**	**t. value**
Participants							
Female participants V Male participants – Curls Still	**Female**	**Male**	**Female**	**Male**	39	.45	2.02
	20.75	15.66	25.98	16.94			
Female participants V Males participants – Curls Dynamic	**Female**	**Male**	**Female**	**Male**	40	.32	2.02
	10.36	14.43	7.46	15.88			

Female participants V Male participants – Macaw Still	**Female**	**Male**	**Female**	**Male**	40	.43	2.02
	18.38	14.17	19.64	15.12			
Female participants V Male participants – Macaw Dynamic	**Female**	**Male**	**Female**	**Male**	40	.39	2.02
	17.90	14.24	13.20	13.83			
Female participants V Male participants – Jack Still	**Female**	**Male**	**Female**	**Male**	40	.11	2.02
	19.60	11.48	21.63	10.13			
Female participants V Male participants – Jack Dynamic	**Female**	**Male**	**Female**	**Male**	40	.04 *	2.02
	25.92	12.33	29.72	11.36			

Note: * indicates statistical significance at 5% level, i.e. a p.value less than .050

Finally, all results of the tests between avatar genders for the emotionally expressive avatars were not statistically significant.

4.5.3 Godspeed Results – Gender Differences

When testing for differences between participants' genders on the data gathered from the Godspeed survey, the results returned no statistically significant differences. All means, standard deviations and *p*. values for these tests can be found in Appendix D. In the final section of this results chapter, the combined results of the EMG and Godspeed data will be discussed.

4.6 Correlation Analysis of EMG and Godspeed Data

A series of correlations were performed on data gathered from EMG and Godspeed implements in order to determine if there were any correlations. Firstly correlations between the individual measures of ‘humanness’, ‘eeriness’ and ‘attractiveness’ from the Godspeed survey were considered, followed by an analysis of correlations between the two (EMG and Godspeed) data sets. The identified correlations were then analysed for statistical significance. The Godspeed was not administered in the baseline section of the experiments, therefore this section will report the results gathered from the fidelity and emotional expression avatars. All correlation matrixes can be found in Appendix E.

4.6.1 Fidelity Avatars

In regards to Ira displaying a happy expression, there was a strong correlation between the ‘humanness’ and ‘attractiveness’ Godspeed indices r (19) = .81. However, this correlation was not statistically significant. See Table 22, Appendix E. In contrast, for the sad expression there was a correlation between ‘humanness’ and ‘attractiveness’ that was statistically significant $r(19) = .70$, $p < .001$. See Table 23, Appendix E.

The results for Emily indicate that there is a correlation between the Godspeed indices of ‘humanness’ and attractiveness for both the happy $r(19) = .64$, $p < .001$ and sad $r(19) = .74$, $p < .001$ avatars, both of which are statistically significant. See Table 24 and Table 25 in Appendix E.

For Liliwen, there were no correlations found between the EMG and Godspeed data, or the within the Godspeed data itself. See Table 26 and Table 27 in Appendix E. In contrast, the results for Leo were slightly correlated in regards to the Godspeed indices of ‘humanness’ and ‘attractiveness’ for happy $r(19) = .69$, $p = .004$ and sad $r(19) = .55$, $p = .009$. Both of the results for Leo are statistically significant, see Table 28 and Table 29 in Appendix E.

4.7 Emotional Expression Avatars – Still Images

The results for the still image of Curls displaying a happy expression has a slight correlation between the Godspeed indices of 'humanness' and 'attractiveness' that is statistically significant $r(19) = .67$, $p = .008$. Similarly, there is a slight correlation between the Godspeed indices of 'humanness' and 'attractiveness' that is statistically significant for Curls displaying a sad expression $r(19) = .56$, $p = .007$. In addition, there is a slight correlation between the Curls sad expression EMG data and the Godspeed indices 'eeriness' that is statistically significant $r(19) = .64$, $p < .001$. See Table 30 and Table 31 in Appendix E

In contrast, there are no correlations between the EMG and Godspeed data for the happy and sad still images for the Jacqueline avatars. These results can be found in Table 32 and Table 33 Appendix E.

Finally, the results for Macaw displaying a happy expression in a still image show that there is a slight correlation between the Godspeed indices of 'humanness' and 'attractiveness' that is statistically significant $r(19) = .50$, $p = .002$. There is a similar correlation in the still image of Macaw displaying a sad expression that is statistically significant $r(19) = .69$, $p = .004$. See Table 34 and Table 35 in Appendix E.

4.8 Emotional Expression Avatars – Animated

The results for the animated versions of Curls displaying a happy expression show that there is a correlation between the Godspeed indices of 'humanness' and 'attractiveness' that is statistically significant $r(19) = .70$, $p = .004$. Similarly, there is slight correlation between the Godspeed indices of 'humanness' and 'attractiveness' that is statistically significant for the sad expression $r(19) = .68$, $p = .006$. See Table 36 and Table 37 in Appendix E.

In contrast to the still version of Macaw happy, the animated version has no correlations. However, there is a slight correlation between the Godspeed indices of 'humanness' and 'attractiveness' that is statistically significant for the sad expression displayed by Macaw $r(19) = .64$, $p < .001$. See Table 38 and Table 39 in Appendix E.

Finally, similar to the still images Jacqueline there are no correlations between the EMG and Godspeed data for the happy and sad still animations. See Table 40 and Table 41 in Appendix E. The final discussion of this section will summarise the details of this chapter.

4.9 Summary of Results

The contents of this chapter have detailed the results of the experiments undertaken for this research. The experiment generated three data sets that were reported in this chapter. The first data set related to the demographic questionnaire administered to each participant prior to the experiment commencing. This data set gave a description of the sample selected for this study in relation to gender, age, education and details how often and on what media participants interacted with avatars.

The second data set that was discussed included the data gathered from the participants startle response, when viewing the baseline images, fidelity and emotional expression avatars. The EMG data gathered was analysed using a series of T-Tests in order to determine if there were any statistically significant differences in the mean scores, for any of these sections in the experiment.

The third data set evaluated in this section related to the Godspeed survey that was administered to each participant for every avatar. This survey allowed participants to rate each avatar's perceived 'humanness', 'eeriness' and 'attractiveness'. The data gathered from this survey was analysed through a repeated measures Analysis of Variance (ANVOA) to test

for statistically significant differences in the participants' perception of the avatars presented to them.

An additional set of tests were carried out to determine if there were any gender differences relating to the participants responses or between the avatars gender. Both of these tests were performed using the EMG and Godspeed data. While there were some statistically significant differences found in the EMG data set, there were no statistically significant differences found in the Godspeed data set.

Finally, the EMG and Godspeed data sets were tested for correlations for statistically significant differences in the results of each participants' responses to EMG and Godspeed testing implements. This indicated whether or not there were any correlations between the responses given for the EMG and Godspeed data sets. These correlations were also tested for statistically significant differences through a series of regressions in order to generate a p. value, to indicate whether or not a correlation was statistically significant.

In the final chapter of this thesis the results of the experiments will be summarised in full. The contributions of this research, as well as a discussion of the limitations of this work and future work will also be outlined in the following section.

5 DISCUSSION AND CONCLUSIONS

5.1 Introduction

Traditionally considered a mainstay of game based interactive entertainment, avatars fulfil a number of functions such as a tool for self-expression and a mechanism for improving story-based interaction. However, the use of digitally rendered characters in films, and the increasing use of avatars in serious game applications, raises questions about how accurately, and cost-effectively, avatars can be used to convey human emotional expressions. This is of particular interest for the area of subject matter driven simulation training, where real human actors communicate through avatars for training purposes. Rapid, cost effective avatar generation tools that use markerless motion capture technology are attractive for these purposes. However, the suitability of avatars generated using this technology to communicate emotionally charged scenarios is not well understood, with the concern that avatar *uncanniness* may occur and impact on training outcomes.

The concept of the Uncanny Valley explains how avatars that are close, but not exactly, humanlike in appearance may produce a level of *uncanniness* that makes them unpleasant to view. In order to extend the existing knowledge surrounding avatars and their levels of fidelity and emotional expression in relation to the Uncanny Valley, this thesis aimed to answer the following questions:

Research Question 1: ***How does avatar fidelity or realism influence the emotional experience of the interactions between humans and computer-generated avatars?***

Research Question 2: ***How do the emotional expressions of avatar facial features affect participants in relation to the levels of valence experienced?***

In order to answer these research questions, a mixture of experimental and survey methods were used, incorporating a three part experiment design. Firstly, participants viewed

a set of nine (9) images from the International Affective Picture System (IAPS) in order to evaluate their individual reactions to different types of images. These images were selected based on their valence rating, and were arranged into the categories of *pleasant*, *neutral* and *unpleasant*, in order to compare participants' reactions with their responses to. For the second part of the experiment, participants viewed a set of high and low fidelity, female and male avatars. In the final part of the experiment, participants viewed a combination of still and animated avatars displaying neutral to happy, and neutral to sad, emotional expressions. For each set of stimuli, participants' startle reflex response to an acoustic white noise probe was recorded. After viewing each of the avatars, participants were asked to complete a Godspeed survey, which asked participants' to rate avatars based on their perceptions on the avatars 'humanness', 'eeriness' and 'attractiveness'. The combination of these implements was used in order to answer the research questions identified by this study.

In answer to research question 1, the level of fidelity or realism's influence on the emotional experience between humans and avatars indicated several interesting findings. Firstly, the higher the level of fidelity that an avatar possessed, the higher the expectation is that the avatar itself will behave in a more human-like manner. The data illustrated that Ira, who is considered a higher fidelity avatar than Emily did not outperform her in regards to 'humanness' or 'attractiveness'. This could indicate that despite the high-fidelity appearance of Ira, his facial expressions were not perceived as positively as Emily's. Secondly, it is worth noting that the digital nature of avatars may also contribute to the level of uncanniness, either through the creation or performance of the avatar. For example, there was a statistically significant difference between Emily and Liliwen. Emily was created and performed through the use of highly specialised equipment, whereas Liliwen was created by off-the-shelf software representing a do-it-yourself type of animation. These differing techniques may potentially contribute to the level of uncanniness and could be further explored in future

work. This could assert as to whether or not an off-the-shelf software package could outperform, or attain the level of, an avatar made similarly to Emily or Ira. However, the data shows that Emily is perceived as most human and most attractive on the Godspeed indices. Thirdly, the results of this study show that the level of 'eeriness' is the same for all levels of fidelity for the avatars utilised in this study, indicating that the avatars produced using low cost and quick methods produced comparative results. Finally, the gender of an avatar emerged as a key influence in human-avatar interaction, as shown by the female avatars outperforming their male counterparts regardless of fidelity level. This could mean that there is a difference in the perception of avatar based on the gender of the avatar, which should be further explored in future work.

In answer to research question 2, the emotional expressions of avatar facial features had an affect on participants' level of valence, with noteworthy findings. Firstly, and again relating to gender, there may be a case for using avatars of different genders to elicit emotional responses from users, which would need to be explored further in future research. Secondly, the recorded responses indicated that the emotional experience of interacting with a smiling avatar was significantly more negative than when viewing an avatar with a sad emotional expression. Thirdly, the factors of avatar gender and realism also play a role in the affective experience that participants underwent when confronted with various levels of realism and facial features.

A discussion of key results, and specific contributions of the research, are discussed in the following section.

5.2 Contributions of the Research

5.2.1 Contributions from the Literature Review

From the combination of a systematic and narrative literature review, the literature located for this research contributed three notable findings. Firstly, the articles analysed demonstrated that there is no universally acknowledged definition of fidelity or realism that directly relates to avatars. In addition, there is also no formal scale or hierarchy that indicated how fidelity and realism should be measured. The mixture of these two gaps in the literature could potentially lead to issues surrounding the development of avatars. With no formal definition of fidelity and realism, as well as the lack of a consistent instrument to measure them, the task of creating and comparing avatars of varying levels of realism and/or fidelity is difficult for designers and researchers. While this research did not focus on this task, some contribution has been made through the development of a scheme to categorise avatars in this thesis.

The second contribution from the literature review involves the identification of the lack of a clear definition, and measurement instrument, for the display of avatar emotional expressions. The lack of definition and measurement could present creation issues for avatar designers and performers. For example, if an avatar is attempting to convey an emotional expression in order to communicate with a human user, and even slight variations in the subtlety of the emotional expression they use exists users may experience a negative emotional experience, and the results of different scenarios can not be compared. It is therefore suggested that research aimed at defining and developing a standard for how emotional expressions in avatars are created, and then displayed by an avatar, could lead to better design choices.

Thirdly, this research highlighted the lack of a universal process for avatar creatation. While several techniques were identified in the literature review, it should be noted that this is not a complete listing of all available development mediums for avatars. The methods described however all have their own approach to avatar development and performance. Some of these articles created their own software while others utilised off the shelf software with mixed results. The literature identified in this research has identified that there is no universal solution to the creation and performance of avatars in order for effective human-avatar interaction to take place. More importantly, there is no framework that categorises the various techniques available in a systematic way, allowing developers and designers to evaluate different approaches in relation to cost, time and performance.

5.2.2 Contributions from the Experiments

In order to measure the affective reactions experienced by participants in relation to pleasant/unpleasant visual stimuli, this study made use of the International Affective Picture System (IAPS). Results for baseline EMG responses to a set of standardised images produced mixed results, with those from male participants similar to those expected from the literature. For all participants, there was a strong response to the pleasant images, however, the responses for neutral and unpleasant images were the reverse of what was expected from the relevant literature (Cuthbert, Bradley, & Lang, 1996; Lang et al., 1998; Vrana et al., 1988). Females consistently recorded a stronger response for all three categories of pleasant, unpleasant and neutral in comparison to their male counterparts. Males tended to have a stronger response for unpleasant in contrast to the responses for the pleasant imagery presented to them. However, the differences between the genders were not statistically significant for this research, although potentially a larger sample size could alter this result.

The evident gender differences do however highlight the need to consider this aspect when designing avatars, and also suggests possible directions for future research.

The EMG startle reflex results relating to the fidelity avatars indicated that the responses were generally higher for both female avatars (i.e. Emily and Liliwen). VanOyen Witvliet and Vrana (1995) state that higher responses are associated with lower arousal and more negative valence (intrinsic attractiveness). This could potentially indicate that these avatars have a lower level of pleasantness that users associate with them during their interactions, which could impact on emotional experience.

From the fidelity experiments, participants were observed to have more potentiated EMG responses to high fidelity sad expressions in the female avatars, and happy expressions in the male avatar. This suggests that different avatar genders may be more suitable for arousing emotional responses in viewers, depending on the emotion to be expressed. Regardless of avatar realism, avatars with happy expressions elicit higher responses from participants than sad expressions; a somewhat surprising result that suggests happy emotions may be more difficult to convey via avatars. This result does however receive support from the literature in which difficulties obtaining the fine muscular movements associated with smiling in avatars are recognised as leading to perceptions of uncanniness (Ochs, Prepin & Pelachaud, 2013; Moser, Derntl, Robinson, Fink, Gur, & Grammer, 2007; Wang & Geiger, 2011).

The EMG data gathered from part three of the experiments demonstrated that there is complex relationship between the emotional expression of an avatar and their level of realism. This is most evident with the low realism ‘Curls’ avatar, when displaying a sad expression, obtaining the second lowest response. This suggests that the emotional expression, and the avatar used to convey it, has a substantial impact on participant perceptions, which becomes most apparent with low levels of realism. However, it should be

noted that overall, the happy emotional expressions recorded a higher response than sad emotional expressions. This suggests that the act of smiling through these avatars may be perceived as more unpleasant than frowning, which again could be attributed to the lack of proper movement around the eyes or a lack of expressive wrinkling as suggested by Courgeon, Buisine and Martin (2009). However, as discussed by Ochs, Prepin, and Pelachaud (2013), who argue that smiling is an important social cue that can communicate emotions as well as stances, these cues are hugely influential in communication and interaction between humans, as well as humans and human-like entities. Therefore, extreme care should be taken when performing smiles through virtual-humans, in order for the appropriate cues related to the intended emotions and stances to be portrayed in such a way that the emotional experience is not unpleasant for the viewer. Finally, from the EMG data, the low response attained by avatars being displayed through a still image and low realism suggests that may not be as effective at communicating emotional expressions as their animated counterparts.

The use of the Godspeed instrument in the experiments also yielded some important contributions. The fidelity, realism and emotional expression of avatars impacts of participants' perceptions of the 'humanness' and 'attractiveness' of an avatar, but not the 'eeriness'. These results suggest that the concept of *uncanniness*, which is closely related to 'eeriness', is somewhat fixed to the digital nature of avatars, rather than specific aspects of the way they are enacted. This is an important finding as it implies that both high and low cost avatar creation techniques can achieve results that are comparable on this dimension.

The use of the modified Godspeed index also provided useful insights on how avatars of different genders are perceived. Godspeed mean values for the perception of 'humanness' in relation to fidelity avatars showed that Emily is perceived as more human than her male

counterpart, Ira. Interesting this occurs for both happy and sad expressions, thus Emily consistently outperforms Ira in regards to perceived 'humanness'.

According to Paul Debevec, project leader for Digital Ira (USC ICT, 2015), Emily was created in an offline scenario, with her face only being displayed from the front, with no close-ups. In contrast, Ira was designed to have his face displayed at all angles including close-ups performed in real time. Both Emily and Ira were created by the University of Southern California, however Ira is technically a higher fidelity than Emily as he was created some years after Emily. It can therefore be assumed that the differences relate to the gender of the avatar, or the likeability/attractiveness of the avatar, rather than the level of fidelity the avatars are associated with. This could relate to some of the key dimensions of uncanniness as described in the literature review by MacDorman, Green, Ho and Kock (2009). From this work, the concept of evolutionary aesthetics, that is a human's ability to locate an appropriate mate, could be effecting the judgement of the user. If a user finds an avatar unlikeable and unattractive, they may experience a negative emotional experience and therefore attempt to avoid the interaction as a form of threat avoidance. It is interesting to note that the fidelity avatars that displayed a sad expression were perceived as more human as opposed to their happy counterparts.

The perception of 'humanness' within the Godspeed survey for the set of emotional expression avatars revealed that the male 'Macaw' avatar was considered to be the most human. This set of avatars consisted of two females and one male, and in contrast to the fidelity avatars where a female was considered the most human, the only male of this set is considered the most human. Despite the classification of 'Curls' being low-realism, she is perceived as more human than the mid-realism avatar of 'Jacqueline'. This suggests that the level of realism does has an impact on the perception of 'humanness' of an avatar, however the results are counter intuitive given the low-realism, cartoonish avatar was considered more

human than the mid-realism avatar generated from a photograph of an actual human. However, the Curls and Macaw avatars were both created and performed using the Faceshift software, whereas Jacqueline was created using the 3D Avatar Store. While both Macaw and Jacqueline have the same level of realism, the techniques for creation and performance of these avatars should be taken into consideration when considering the perception of 'humanness' of these avatars.

Interestingly, from the Godspeed survey, the perception of 'attractiveness' indicates that Curls is the most attractive of all avatars within this set. This could be due to the low-realism of Curls, who could be identified by the participants as a low-threat (see MacDorman, Green, Ho and Kock (2009). This low-threat is associated with a lack of harm to the users, and of a lack threat to their *unique* humanness, and therefore participants may find Curls more attractive than the other two avatars.

Interestingly, for both the fidelity and emotional expression sets of avatars, there were no statistically significant differences in the level of 'eeriness'. This contradicts current research that associates avatars and the Uncanny Valley with high levels of 'eeriness'. Future work could be conducted on this anomaly to determine if this was due to the sample size or due other factors with avatars and the uncanny valley.

Finally, a series of correlations were run in order to determine whether or not there were correlations between the EMG and Godspeed data sets. The outcome of these tests revealed that there was only one instance of a significant correlation. The majority of the correlations came from within the Godspeed indices, specifically between the 'humanness' and 'attractiveness' indices, neither of which were paired with 'eeriness', which matches the relevant literature. Ho and MacDorman (2010) argue that the indices of 'humanness' and 'attractiveness' are not significantly correlated with 'eeriness'. However, the independence related to the 'eeriness' and 'humanness' indices indicate that they could act as the x and y

axis of the original Uncanny Valley graph, allowing digital anthropomorphic characters to be plotted in Mori's (1970) graph.

In the final section of this thesis, the limitations of the study as well as avenues for future work will be discussed.

5.3 Limitations and Future Work

While some variation was expected within the baseline EMG responses, the responses gathered from one participant were considered outside of the normal range and thus were removed from the final analysis. However, the remaining responses were within the expectations of this study. The literature associated with the IAPS imagery does refer to potential differences in female and male responses (*International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8.* Lang, Bradley, & Cuthbert, 2008.p. 28-42 & 43-57). However, it is not made clear in this literature if these differences are statistically significant. The results from this research indicate that gender differences may extend beyond the level response and impact on the perception of images themselves. Further future work could include, a better understanding of the gender difference in obtaining responses when viewing standardised imagery such as the International Affective Picture System (IAPS), which could be applied to understanding how different genders respond to virtual-humans. In addition, the level of fidelity and realism in relation to emotional expressions could be explored further with a larger sample size in order to test for differences in the perception of avatars.

The startle reflex EMG signals were not able to produce any statistically significant results. A potential reason for non-significant EMG T-Test results could be due to the small sample size, and the inherently large variability found in human EMG startle responses (Blumenthal et al., 2005). Due to the nature of an Honours project, it is difficult to gather

large numbers of participants. This is primarily due to time restrictions of the overall duration of an Honours course. However, the methods of this research are designed to be replicable in future studies. Regardless, these preliminary results appear to indicate some differences in the emotional response of participants to avatar emotional expressions.

In regards to the stimuli presented during experiments, a photograph or video of a 'real' human being was not used. This could potentially be included in future studies to set a reference point for findings, and also to determine how accurately participants can identify stimuli as either photo-realistic avatars or real human beings. The choice of 'real' human being would need to be carefully researched in order to ensure that no unintentional bias was introduced into the recording and experiment.

Finally, the literature revealed that there are several techniques that could improve the fidelity of an avatar. It is thus noted that being able to produce realistic colouring of avatar faces would add another of level to the fidelity in an avatar, which could allow for a greater acceptance of a computer generated avatar. This, together with exploration of cultural issues associated with avatars, are interesting and valid directions for future research.

5.4 Conclusion

This research represents an entry point into what is a broad, cross-disciplinary research area. While a number of important findings and contributions have been made, the work also addresses the fundamental research tenet of posing as many, if not more, questions than it answers.

Additionally, a significant amount of data has been captured during the experiments. While analysis of data necessary to answer the research questions has been presented in this thesis, some additional analysis will be undertaken and published outside of the thesis via a conference publication.

6 REFERENCES

Alexander, A. L., Brunyé, T., Sidman, J., & Weil, S. A. (2005). From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in pc-based simulations and games. *DARWARS Training Impact Group*, *5*, 1-14.

Alkawaz, M. H., & Basori, A. H. (2012, December). The effect of emotional colour on creating realistic expression of avatar. In *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry* (pp. 143-152). ACM.

Annetta, L. A., Murray, M. R., Laird, S. G., Bohr, S. C., & Park, J. C. (2006). Serious games: Incorporating video games in the classroom. *Educause Quarterly*, *29*(3), 16.

Arellano, D., Varona, J., & Perales, F. J. (2008). Generation and visualization of emotional states in virtual characters. *Computer Animation and Virtual Worlds*, *19*(3-4), 259-270.

Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annual review of psychology*, *58*, 373.

Bartholow, B. D., Bushman, B. J., & Sestir, M. A. (2006). Chronic violent video game exposure and desensitization to violence: Behavioral and event-related brain potential data. *Journal of Experimental Social Psychology*, *42*, 532-539.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2008). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, *1*(1), 71-81.

Bernat, E., Patrick, C. J., Benning, S. D., & Tellegen, A. (2006). Effects of picture content and intensity on affective physiological response. *Psychophysiology*, *43*(1), 93-103.

Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V., & Van Boxtel, A. (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology*, *42*(1), 1-15.

Bradley, M. (2014, February 17). The Center for the Study Of Emotion And Attention. Retrieved May 3, 2015, from http://csea.phhp.ufl.edu/media.html

Bradley, M. M., Cuthbert, B. N., & Lang, P. J. (1996). Picture media and emotion: Effects of a sustained affective context. *Psychophysiology*, *33*(6), 662-670.

Brenton, H., Gillies, M., Ballin, D., & Chatting, D. (2005, September). The uncanny valley: does it exist. In *Proceedings of conference of human computer interaction, workshop on human animated character interaction*.

Briner, R. B., & Denyer, D. (2012). Systematic review and evidence synthesis as a practice and scholarship tool. *Handbook of evidence-based management: Companies, classrooms and research*, 112-129.

Burleigh, T. J., Schoenherr, J. R., & 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.

Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *Affective Computing, IEEE Transactions on*, *1*(1), 18-37.

Chiang, C. C., Chen, Z. W., & Yang, C. N. (2013). Video-Driven Creation of Virtual Avatars by Component-based Transferring of Facial Expressions. *Journal of Information Science and Engineering*, *29*(6), 1265-1283.

Cipriani, A., & Geddes, J. (2003). Comparison of systematic and narrative reviews: the example of the atypical antipsychotics. *Epidemiologia e psichiatria sociale*, 12(03), 146-153.

Courgeon, M., Buisine, S., & Martin, J. C. (2009, January). Impact of expressive wrinkles on perception of a virtual character's facial expressions of emotions. In *Intelligent Virtual Agents* (pp. 201-214). Springer Berlin Heidelberg.

Creswell, J. (2007). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage.

Davis, J. & Mosdell, N. (2006) *Practical Research Methods for Media and Cultural Studies: Making people count*. Edinburgh: Edinburgh University Press

De Zwart, M., & Lindsay, D. (2012). My Self, My Avatar, My Rights? Avatar Identity in Social Virtual Worlds. *At The Interface / Probing The Boundaries*, *85*, 81-100.

Dietrich, D. R. (2013). Avatars of Whiteness: Racial expression in video game characters. *Sociological Inquiry*, *83*(1), 82-105.

Elsesser, K., Sartory, G., & Tackenberg, A. (2004). Attention, heart rate, and startle response during exposure to trauma-relevant pictures: a comparison of recent trauma victims and patients with posttraumatic stress disorder. *Journal of abnormal psychology*, *113*(2), 289.

Elson, M., & Quandt, T. (2014). Digital Games in Laboratory Experiments: Controlling a Complex Stimulus Through Modding.

FaceShift Studio (Version 2014.2.01) [Computer Software]. (2014). Retrieved From http://www.faceshift.com/product/

Farbi, M., & Moore, D. (2005). The use of emotionally expressive avatars in Collaborative Virtual Environments. *Virtual Social Agents,* 88.

Fidelity. (2015). In *Oxford Dictionaries*. Retrieved from http://www.oxforddictionaries.com/definition/english/fidelity

Filion, D. L. (1998). *The psychological significance of human startle eyeblink modification: a review*. Biological Psychology

Filion, D. L., Dawson, M. E., & Schell, A. M. (1998). The psychological significance of human startle eyeblink modification: a review. *Biological psychology*, *47*(1), 1-43.

Fontaine, J. R., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological science*, *18*(12), 1050-1057.

Frijda, N. H. (1986). The emotions: Studies in emotion and social interaction. *Paris: Maison de Sciences de l'Homme*.

Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., & Sasse, M. A. (2003, April). The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 529-536). ACM.

Geller, T. (2008). Overcoming the uncanny valley. *IEEE Computer Graphics and Applications*, *28*(4), 11-17.

Goetzman, G. (Producer), Starkey, S. (Producer), Teitler W. (Producer), Zemeckis, R. (Producer) & Zemeckis, R. (Director). (2004). *The Polar Express*. [Motion Picture]. United States: Warner Bros. Pictures

Grimshaw, M., Lindley, C. A., & Nacke, L. E. (2008). Sound and immersion in the first-person shooter: mixed measurement of the player's sonic experience.

Guest, G., Bunce, A. and Johnson, L. (2006), 'How many interviews are enough?' *Field Methods, 18*, 59-82.

Guynup, S., Broglio, R., & Demmers, J. (2004, August). Teaching beyond the human form: avatar as multimedia expression. In *ACM SIGGRAPH 2004 Educators program* (p. 18). ACM.

Harrell, D. F., & Harrell, S. V. (2012). Imagination, Computation, and Self-Expression: Situated Character and Avatar Mediated Identity. *Leonardo Electronic Almanac, 17*(2).

Hess, U., Sabourin, G., & Kleck, R. E. (2007). Postauricular and eyeblink startle responses to facial expressions. *Psychophysiology*, *44*(3), 431-435.

Ho, C. C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, *26*(6), 1508-1518.

IBM SPSS Statistics Standard (Version 23.0) [Computer Software] (2011) Retrieved From: http://www-03.ibm.com/software/products/en/spss-stats-standard

Keele, S. (2007). *Guidelines for performing systematic literature reviews in software engineering* (pp. 1-57). Technical report, EBSE Technical Report EBSE-2007-01.

Kim, S. H. (2007). Generation of expression space for realtime facial expression control of 3D avatar. In *Computer Vision/Computer Graphics Collaboration Techniques* (pp. 318-329). Springer Berlin Heidelberg.

Koda, T., & Ishida, T. (2006, January). Cross-cultural study of avatar expression interpretations. In *Applications and the Internet, 2006. SAINT 2006. International Symposium on* (pp. 7-pp). IEEE.

Labchart (Version 8.0.8) [Computer Software] (2015) Retrieved From: http://www.adinstruments.com/support/downloads/windows/labchart

Lang, P. J. (1995). The emotion probe: studies of motivation and attention. *American psychologist*, *50*(5), 372.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1998). Emotion, motivation, and anxiety: brain mechanisms and psychophysiology. *Biological psychiatry*, *44*(12), 1248-1263.

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8.* University of Florida, Gainesville, FL.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2007). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, 39-58.

Lee, S., Carlson, G., Jones, S., Johnson, A., Leigh, J., & Renambot, L. (2010, January). Designing an expressive avatar of a real person. In *Intelligent Virtual Agents* (pp. 64-76). Springer Berlin Heidelberg.

Li, L., Liu, G., Zhang, M., Pan, Z., & Song, E. (2010). BAAP: a behavioural animation authoring platform for emotion driven 3D virtual characters. In *Entertainment Computing-ICEC 2010* (pp. 350-357). Springer Berlin Heidelberg.

Liu, J., Chen, Y., Gao, X., Xie, J., & Gao, W. (2008). AAML Based Avatar Animation with Personalized Expression for Online Chatting System. In *Advances in Multimedia Information Processing-PCM 2008* (pp. 643-652). Springer Berlin Heidelberg.

Liu, Z. (2006). Emotion interaction of virtual character in 3D artificial society. In *Simulated Evolution and Learning* (pp. 648-655). Springer Berlin Heidelberg.

MacDorman, K. F., Green, R. D., Ho, C. C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in human behavior*, *25*(3), 695-710.

MATLAB and Statistics Toolbox Release R2014b, The MathWorks Inc., Natick, Massachusetts, United States.

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition and emotion*, *23*(2), 209-237.

Microsoft. (2013). Microsoft Excel [computer software]. Redmond, Washington: Microsoft.

Mitchell, W. J., Szerszen Sr, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., & MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception*, *2*(1), 10.

Moser, E., Derntl, B., Robinson, S., Fink, B., Gur, R. C., & Grammer, K. (2007). Amygdala activation at 3T in response to human and avatar facial expressions of emotions. *Journal of neuroscience methods*, *161*(1), 126-133.

Mori, M. (1970**)** *The Uncanny Valley* [Image] Retrieved from: http://spectrum.ieee.org/automaton/robotics/humanoids/the-uncanny-valley

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *Robotics & Automation Magazine, IEEE*, *19*(2), 98-100.

Mukherji, P & Albon, D. (2015). Chapter 1. In *Research Methods in Early Childhood: An Introductory Guide* (2nd ed.). Sage.

Mundy-Castle, A. C., & McKiever, B. L. (1953). The psychophysiological significance of the galvanic skin response. *Journal of Experimental Psychology*, *46*(1), 15.

Noël, S., Dumoulin, S., & Lindgaard, G. (2009). Interpreting human and avatar facial expressions. In *Human-Computer Interaction–INTERACT 2009* (pp. 98-110). Springer Berlin Heidelberg.

Noël, S., Dumoulin, S., Whalen, T., & Stewart, J. (2006). Recognizing emotions on static and animated avatar faces. In *Haptic Audio Visual Environments and their Applications, 2006. HAVE 2006. IEEE International Workshop on* (pp. 99-104). IEEE.

Philipp, M. C., Denney, H. T., Moncrieft, A., Sharma, R., & Vanman, E. J. (2007, January). The Effects of Avatar Presence and Emotional Valence on Facial Muscle Activity in an Immersive Virtual Environment. In *PSYCHOPHYSIOLOGY* (Vol. 44, pp. S39-S39). 9600 Garsington RD, Oxford ox4 2dq, Oxon, England: Blackwell Publishing.

Prepin, K., Ochs, M., & Pelachaud, C. (2013). Beyond backchannels: co-construction of dyadic stance by reciprocal reinforcement of smiles between virtual agents. In *International Conference CogSci (Annual Conference of the Cognitive Science Society)*.

Proulx, T., Heine, S. J., & Vohs, K. D. (2010). When is the unfamiliar the uncanny? Meaning affirmation after exposure to absurdist literature, humor, and art. *Personality and Social Psychology Bulletin*.

Puklavage, C., Pirela, A. R., Gonzalez, A. J., & Georgiopoulos, M. (2010, June). Imitating personalized expressions in an avatar through machine learning. In *Twenty-Third International FLAIRS Conference*.

Ravaja, N., Saari, T., Salminen, M., Laarni, J., & Kallinen, K. (2006). Phasic emotional reactions to video game events: A psychophysiological investigation. *Media Psychology*, *8*(4), 343-367.

Realism. (2015). In *Oxford Dictionaries*. Retrieved from http://www.oxforddictionaries.com/definition/english/realism

Richey, R. C. (1998). The pursuit of useable knowledge in instructional technology. *Educational Technology Research and Development*, *46*(4), 7-22.

Riek, L. D., & Watson, R. N. (2010). The Age of Avatar Realism: When Seeing Shouldn't be Believing. *IEEE robotics & automation magazine*, *17*(4), 37-42.

Robson, C. (2002) *Surveys and questionnaires* Oxford; Malden, Mass. : Blackwell Publishers

Rosen, K. R. (2008). The history of medical simulation. *Journal of critical care*, *23*(2), 157-166.

Scherer, K. R. (1993). Neuroscience projections to current debates in emotion psychology. *Cognition & Emotion*, *7*(1), 1-41.

Seddik, B., Maâmatou, H., Gazzah, S., Chateau, T., & Ben Amara, N. E. (2013, March). Unsupervised facial expressions recognition and avatar reconstruction from Kinect. In *Systems, Signals & Devices (SSD), 2013 10th International Multi-Conference on* (pp. 1-6). IEEE.

Seyama, J. I., & Nagayama, R. S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments*, *16*(4), 337-351.

Steptoe, W., & Steed, A. (2008, March). High-fidelity avatar eye-representation. In *Virtual Reality Conference, 2008. VR'08. IEEE* (pp. 111-114). IEEE.

Tinwell, A. (2014). *The Uncanny Valley in Games and Animation*. CRC Press. Boca Raton, Florida.

Tranfield, D., Denyer, D., & Smart, P. (2003). Toward a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, *14*, 207–222.

USC ICT (2015). *Digital Ira*. Retrieved from http://ict.usc.edu/prototypes/digital-ira/

Vaishnavi, VK. & Kuechler Jr, W. (2007) *Design science research methods and patterns: innovating information and communication technology*

VanOyen Witvliet, C., & Vrana, S. R. (1995). Psychophysiological responses as indices of affective dimensions. *Psychophysiology*, *32*(5), 436-443.

Vaškevičius, E., Bazevičius, E., Vidugirienė, A., & Kaminskas, V. (2013). Investigation of Features for User's Emotional State Control Using Virtual Character. *Procedia-Social and Behavioural Sciences*, *83*, 808-813.

Vasalou, A., & Joinson, A. N. (2009). Me, myself and I: The role of interactional context on self-presentation through avatars. *Computers in Human Behavior*, *25*(2), 510-520.

Vrana, S. R., Spence, E. L., & Lang, P. J. (1988). The startle probe response: a new measure of emotion?. *Journal of abnormal psychology*, *97*(4), 487.

Wang, Y., & Geiger, J. (2011). Using facial emotional signals for communication between emotionally expressive avatars in virtual worlds. In *Affective Computing and Intelligent Interaction* (pp. 297-304). Springer Berlin Heidelberg.

Weise, T., Bouaziz, S., Li, H., & Pauly, M. (2011, August). Realtime performance-based facial animation. In *ACM Transactions on Graphics (TOG)* (Vol. 30, No. 4, p. 77). ACM.

Weyers, P., Mühlberger, A., Hefele, C., & Pauli, P. (2006). Electromyographic responses to static and dynamic avatar emotional facial expressions. *Psychophysiology*, *43*(5), 450-453.

Wilson, M. L., Mackay, W., Chi, E., Bernstein, M., Russell, D., & Thimbleby, H. (2011, May). RepliCHI-CHI should be replicating and validating results more: discuss. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems* (pp. 463-466). ACM.

Wrzesien, M., Rodríguez, A., Rey, B., Alcañiz, M., Baños, R. M., & Vara, M. D. (2015). How the physical similarity of avatars can influence the learning of emotion regulation strategies in teenagers. *Computers in Human Behaviour*, *43*, 101-111.

Yang, S., & Bhanu, B. (2012). Understanding discrete facial expressions in video using an emotion avatar image. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, *42*(4), 980-992.

Yun, C., Deng, Z., & Hiscock, M. (2009). Can local avatars satisfy a global audience? A case study of high-fidelity 3D facial avatar animation in subject identification and emotion perception by US and international groups. *Computers in Entertainment (CIE)*, *7*(2), 21.

Zalewski, L., & Gong, S. (2005, June). 2d statistical models of facial expressions for realistic 3d avatar animation. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (Vol. 2, pp. 217-222). IEEE.

Ziegeler, D., & Zuehlke, D. (2005, July). Emotional user interfaces and humanoid avatars in industrial environments. In *Proceedings of the 16th IFAC World Congress*.

7 APPENDIX A – EXPERIMENT PRE-QUESTIONNAIRE

ID: _____

Date: ____/____/____

Exploring emotional expression in avatar facial features in relation to the Uncanny Valley

Jacqueline Bailey
jacqueline.bailey@uon.edu.au

Dr Karen Blackmore
School of Design, Communication and Information Technology
Karen.Blackmore@newcastle.edu.au
+61 2 492 15206

Initial document version 2

Welcome to the Emotional Expression in Avatars experiment.

This project is designed to explore the effects that facial movements of virtual characters (avatars) have on observers. In order for effective communication to occur observers must be able to relate to the avatars emotional expression and know how to respond appropriately. Equally, avatars must be able to display emotions that are appropriate to the context that surrounds them which can lead to an effective non-verbal communication between the avatar and the observer.

We have carefully chosen you to participate in this study as you represent a section of the population demographic which will provide us with a balanced view of how different people react to variations of emotional expression in avatars. We would like your feedback on when you felt the any sense of uneasiness when interacting with the avatars emotional expression.

You may feel concerned about being monitored by the startle probe, but rest assured, all information-gathering tools are non-invasive, and you can opt out at any time if you feel uncomfortable.

The researcher will be present during the experiment, but she is only there if you wish to pause or stop the experiment for any reason. All data used in the experiment will be recorded by various apparatus present during the experiment.

Before the experiment begins, we would like you to answer some basic questions about yourself in regards to your demographics. This will help us understand a little bit more about you and the data you will be helping us collect. As you progress through the survey feel free to add any extra information or comments you feel would be helpful. If you do not understand anything on this questionnaire, please just ask and the researcher will be happy to clarify.

Questions begin on the next page.

Questions *(Please circle relevant responses)*

1. **Please indicate gender:**
    - **a.** Female
    - **b.** Male

2. **What is your age range:**
    - **a.** 18-25
    - **b.** 26-40
    - **c.** 41-60

3. **What is your highest level of education qualification attained?**
    - **a.** School Certificate
    - **b.** TAFE
    - **c.** Higher School Certificate
    - **d.** Undergraduate
    - **e.** Postgraduate
    - **f.** Other (please specify)

……

4. **If relevant, what degree are you currently studying?**

……

5. **Where do you interact with virtual characters (avatars)?** *(You can circle as many as needed)*
    - **a.** Movies
    - **b.** Games
    - **c.** Television shows
    - **d.** Web shows
    - **e.** Other (please specify)

……

6. **What platform/s do you interact with these virtual characters (avatars) on?** *(You can circle as many as needed)*
    - **a.** PC
    - **b.** Mac
    - **c.** Mobile device
    - **d.** Television
    - **e.** Other (please specify)

……

7. **On average how many hours a week would you spend interacting any form of media that contains virtual characters (avatars)**
    - **a.** 0 hours a week
    - **b.** 1-5 hours a week
    - **c.** 6-10 hours a week
    - **d.** 11-15 hours a week
    - **e.** 16-20 hours a week
    - **f.** 20-25 hours a week
    - **g.** 25+ hours a week

8. **Do you have any character animation experience?**
    - **a.** None at all
    - **b.** Some experience
    - **c.** Large amounts of experience

9. **Are you familiar with the uncanny valley?**
    - **a.** Not at all
    - **b.** Somewhat
    - **c.** I understand the concept

Finished - Thank you for completing this pre-experiment questionnaire. Please return it to the researcher after completion

8 APPENDIX B – EXPERIMENT GODSPEED SURVEY

Avatar: HF-M-H

Humanness - Please rate your impression of the avatar on these scales:

Artificial	1	2	3	4	5	Natural
Human-made	1	2	3	4	5	Humanlike
Without definite Life-span	1	2	3	4	5	Mortal
Inanimate	1	2	3	4	5	Living
Mechanical movement	1	2	3	4	5	Biological movement
Synthetic	1	2	3	4	5	Real

Eeriness - Please rate your impression of the avatar on these scales:

Reassuring	1	2	3	4	5	Eerie
Numbing	1	2	3	4	5	Freaky
Ordinary	1	2	3	4	5	Supernatural
Bland	1	2	3	4	5	Uncanny
Unemotional	1	2	3	4	5	Hair-raising
Uninspiring	1	2	3	4	5	Spine-tingling
Predictable	1	2	3	4	5	Thrilling
Boring	1	2	3	4	5	Shocking

Attractiveness - Please rate your impression of the avatar on these scales:

Unattractive	1	2	3	4	5	Attractive
Repulsive	1	2	3	4	5	Agreeable
Ugly	1	2	3	4	5	Beautiful
Messy	1	2	3	4	5	Sleek
Crude	1	2	3	4	5	Stylish

9 APPENDIX C – RISK ASSESSMENT FOR THIS RESEARCH

9.1 Risk Identification

Sources of Risk	Type of Risk
Technical	• Loss of data including: o Emails o Data collection samples from the experiments o Including talk aloud if applicable o Written sections of the Honours thesis o Surveys from the experiment if applicable o Literature found o Reference and bibliographic details o Completed pre-experiment surveys o Video recordings if applicable • Startle Probe malfunction o Not recording data o Syncing issues • FaceShift malfunction o Not recording the movements o Not displaying movements as required • Lack of footage from the University of Southern California
External	• ADC delays • Ethics delays

	• Ability to obtain an adequate amount of participants for the experiment • Unable to get to places required by the project due to weather or other circumstances

Organisational	• Resource availability
	  o Other people using the startle probe for their projects
	  o Making the test for participants to view
	  o Making the models in FaceShift for participants to view (both stationary and dynamic)
	• Project Dependences
	  o Ethics approval
	  o ADC approval

Project Management	• Planning
	  o Need a really good and focused schedule
	  o Must have good alternates in case of failing to adhere to schedule due to unforeseen consequences
	• Must adhere to time oriented goals and requirements, including:
	  o Milestones
	  o Meetings
	  o Progress reports
	• Communication
	  o Karen may be busier than usual at times during the semesters, partially if it is exam time or an assignment is due, which may lead to delayed communicating
	  o ADC apparently take a while to authorize things that are needed

	o Gathering people to participate in the tests how to get a decent number of a variety of people to participate?

9.2 Risk Assessment – Scale 1 low 3 Medium 5 High

Risk Event	*Likelihood*	*Impact*	*Detection difficulty*	*When*
Technical				
Loss of data	3	5	1	All stages of the project
Malfunctioning Startle Probe	3	5	1	During test set up or during tests
Malfunctioning FaceShift	3	5	1	When capturing the animations or storage and/or usage of the animations or static imagery
Lack of footage from the University of Southern California	3	3	1	During pre-testing and running of the experiments
External				
ADC delays	4	2	2	All stages of the project
Ethics delays	5	5	1	All stages of the project
Unable to get to university or other location required for the project	3	3	1	All stages of the project
Unable to get enough participants	2	4	2	During the testing stages of the project
Organisational				
Startle probe availability	3	3	1	During the testing stages of the project
Making the tests for participants to view	2	4	2	Before the testing begins
Creating the models both dynamic and static for the	3	4	2	Before the testing begins

participants to view				
Project Management				
Not adhering to schedule	5	5	1	Have alternate plans in order to meet milestones and deadlines
Being able to contact/get responses from Karen	3	3	2	All stages of the project that overlap with times of high student demand
Being able to contact and get a response from the ADC	3	3	2	All stages of the project
OH & S				
Strain from extended computer use	5	5	1	All stages of the project
Personal safety during recruitment of participants	1	5	3	Before the experiment begins
Personal safety during experiments	1	5	3	During experiments

9.3 Response Development

Risk Event	*Response*	*Contingency Plan*	*Trigger*	*Responsibility*
Technical				
Loss of data	Avoid: Have hard copy back-ups, as well as digital on a USB storage device and/or cloud server	Regain what data you can, have a log of your data and where it is stored	Loss of any form of data related to the project	Jacqueline
Malfunctioning Startle Probe	Mitigate: reduce the likelihood of malfunction occurring	Have a subject matter expert on hand in order to understand the how to fix the problem	Startle probe is not recording data or functioning correctly	Jacqueline
Malfunctioning FaceShift	Mitigate: reduce the likelihood of malfunction occurring	Have a subject matter expert on hand in order to understand the how to fix the problem or have a good understanding of the program to begin with	FaceShift is not recording data or functioning correctly	Jacqueline
Lack of footage from the University of Southern California	Mitigate: Source footage of Emily and Ira from the internet	Edit the retrieved footage to the appropriate length for the experiments	No footage supplied, or unusable footage supplied	Jacqueline
External				
ADC delays	Mitigate: reduce the impact that this will	Work on other aspects until delay is resolved	A delay in information or other component required	ADC and Jacqueline

	have on the project		from ADC	
Ethics delays	Mitigate: reduce the impact that this will have on the project	Work on other aspects until delay is resolved	A delay in information or other component required from Ethics	Ethics Committee and Jacqueline
Unable to get to university or other location required for the project	Avoid: be at the appropriate location	Set an alternate time and date to fulfil obligations that suit all parties	Unable to attend campus or other location required for the project due to unforeseen complications	Jacqueline
Unable to get enough participants	Mitigate: reduce the likelihood of not having enough participants	Find an alternate way to gather participants	Not enough participants for testing	Jacqueline
Organisational				
Startle probe availability	Mitigate: reduce the likelihood of not being able to access this equipment by working around other users of the startle probe equipment	Find an alternate time that suits all parties	Not able to access the equipment due to other persons using the startle probe for their testing	Jacqueline
Creating the models both dynamic and static for the participants to view	Mitigate: reduce the likelihood of not having a full functioning models	Find alternate sources of models	Models are not ready for the test	Jacqueline
Project Management				
Not adhering to schedule	Mitigate: reduce the impact that this will have on the project	Have alternate plans ready for when this happens	Schedule not being adhered to	Jacqueline
Being unable to contact/get responses from Karen	Retain: Await response and be aware of busy time periods	Continue with other work while UON Supervisor is unable to be reached	Unable to contact or have responses from UON Supervisor	Jacqueline
Being unable to contact and get a response from the ADC	Retain: Await response	Continue with other work	Unable to contact or get a response for the ADC	Jacqueline
OH & S				
Strain from extended computer use	Mitigate: Do not sit at the computer for more than 2 hours without a stretch	Continue with reading or other non-computer tasks	Arm, back or eye strain	Jacqueline
Personal safety during recruitment of participants	Mitigate: Use only posters or email to distribute recruitment information	Seek assistance from UON supervisor to recruit participants	Not following approved ethics process	Jacqueline

Personal safety during experiments	Mitigate: Conduct experiments during standard office hours or liaise with UON Supervisor for outside of normal office hours	Ensure staff are in nearby offices when conducting experiments; terminate experiment	Feeling unsafe or unsure during an experiment	Jacqueline

10 APPENDIX D – GENDER DIFFERENCES IN GODSPEED MEANS

Table 20 – Female Godspeed Means

Avatar	Humanness			Eeriness			Attractiveness		
	Mean	Std. Deviation	p. value	Mean	Std. Deviation	p. value	Mean	Std. Deviation	p. value
High-fidelity Male	2.94	1.12	.72	2.63	.46	.90	2.76	.85	.68
High-fidelity Female	4.44	.61	.90	2.27	.60	.64	4.34	.63	.17
Low-fidelity Male	1.46	.47	.17	2.40	.55	.19	2.42	.96	.64
Low-fidelity Female	1.71	.71	.37	2.31	.58	.33	3.02	.67	.80
Curls – Still	1.37	.41	.13	2.44	.42	.16	2.93	1.03	.43
Macaw – Still	1.91	.90	.85	2.59	.90	.64	2.25	.72	.90
Jacqueline - Still	1.37	.38	.78	2.29	.58	.18	1.96	.98	.62
Curls – Animated	1.68	.56	.26	2.65	.48	.42	2.92	1.25	.72
Macaw – Animated	1.96	.88	.26	2.84	.66	.99	2.43	.77	.88
Jacqueline - Animated	1.60	.45	.54	2.25	.45	.12	2.24	1.06	.81

Table 21 – Male Godspeed Means

Avatar	Humanness			Eeriness			Attractiveness		
	Mean	**Std. Deviation**	**p. value**	**Mean**	**Std. Deviation**	**p. value**	**Mean**	**Std. Deviation**	**p. value**
High-fidelity Male	2.77	.97	.72	2.66	.40	.90	2.61	.78	.68
High-fidelity Female	4.40	.69	.90	2.39	.46	.64	3.91	.73	.17
Low-fidelity Male	1.81	.69	.17	2.79	.77	.19	2.23	.85	.64
Low-fidelity Female	1.99	.68	.37	2.59	.72	.33	2.93	.95	.80
Curls – Still	1.78	.71	.13	2.87	.84	.16	2.57	.95	.43
Macaw – Still	1.98	.82	.85	2.78	.99	.64	2.21	.75	.90
Jacqueline - Still	1.32	.31	.78	2.72	.84	.18	1.78	.55	.62
Curls – Animated	2.04	.82	.26	2.90	.86	.42	2.72	1.12	.72
Macaw – Animated	2.40	.82	.26	2.84	.90	.99	2.48	.78	.88
Jacqueline - Animated	1.73	.54	.54	2.67	.67	.12	2.14	.77	.81

11 APPENDIX E – CORRELATION MATRIXES FOR ALL AVATARS

11.1 High-Fidelity Avatars

Table 22 – Ira (Happy)

Ira Happy	EMG	Humanness	Eeriness	Attractiveness
EMG	1			
Humanness	0.19413256	1		
Eeriness	-0.2758228	0.16738925	1	
Attractiveness	0.14609009	**0.81515038**	0.09190869	1

Table 23 – Ira (Sad)

Ira Happy	EMG	Humanness	Eeriness	Attractiveness
EMG	1			
Humanness	0.032794481	1		
Eeriness	-0.35029197	-0.15237869	1	
Attractiveness	-0.05941817	**0.701057245**	-0.25714809	1

Table 24 – Emily (Happy)

Emily Happy	EMG	Humanness	Eeriness	Attractiveness
EMG	1			
Humanness	0.398063528	1		
Eeriness	-0.007057283	-0.272279716	1	

Table 25 – Emily (Sad)

Emily Sad	EMG	Humanness	Eeriness	Attractiveness
EMG	1			
Humanness	0.28058449	1		
Eeriness	-0.36795593	-0.03838101	1	
Attractiveness	-0.11666217	**0.74186344**	-0.15591913	1

11.2 Low-Fidelity Avatars

Table 26 – Liliwen (Happy)

Low-fidelity Female Happy	EMG	Humanness	Eeriness	Attractiveness
EMG	1			
Humanness	-0.086852773	1		
Eeriness	-0.444133847	0.166415479	1	
Attractiveness	0.489095981	0.114993705	0.161442568	1

Table 27 – Liliwen (Sad)

Low-fidelity Female Sad	EMG	Humanness	Eeriness	Attractiveness
EMG	1			
Humanness	-0.132843032	1		
Eeriness	-0.088238213	0.127701189	1	
Attractiveness	-0.192688204	0.213058881	0.173742297	1

Table 28 – Leo (Happy)

Low-fidelity Male Happy	EMG	Humanness	Eeriness	Attractiveness
EMG	1			
Humanness	-0.368499492	1		
Eeriness	-0.142123325	0.224098181	1	
Attractiveness	-0.225778534	**0.699206272**	0.348113363	1

Table 29 – Leo (Sad)

Low-fidelity Male Sad	EMG	Humanness	Eeriness	Attractiveness
EMG	1			
Humanness	-0.187177505	1		
Eeriness	0.138064177	0.081884868	1	
Attractiveness	-0.014228752	**0.551210218**	0.108487306	1

11.3 Emotional Expression Avatars – Still Images

Table 30 - Curls (Happy)

Curls Happy	EMG	Humanness	Eeriness	Attractiveness
EMG	1			
Humanness	0.17	1		
Eeriness	0.16	0.48	1	
Attractiveness	0.40	**0.67**	0.24	1

Table 31 – Curls (Sad)

Curls Sad	EMG	Humanness	Eeriness	Attractiveness
EMG	1			
Humanness	0.03	1		
Eeriness	**0.57**	0.21	1	
Attractiveness	0.01	**0.64**	0.17	1

Table 32 – Jacqueline (Happy)

Jacqueline Happy	EMG	Humanness	Eeriness	Attractiveness
EMG	1			
Humanness	-0.02447323	1		
Eeriness	-0.33026091	0.03974116	1	
Attractiveness	0.22431019	0.18290146	0.17022097	1

Table 33 – Jacqueline (Sad)

Jacqueline Sad	EMG	Humanness	Eeriness	Attractiveness
EMG	1			
Humanness	-0.15874922	1		
Eeriness	-0.198545701	0.036004934	1	
Attractiveness	-0.143034126	0.285314826	0.156916841	1

Table 34 - Macaw (Happy)

Macaw Happy	EMG	Humanness	Eeriness	Attractiveness
EMG	1			
Humanness	-0.114887242	1		
Eeriness	-0.485495861	-0.211337938	1	
Attractiveness	0.001246236	**0.500820285**	0.123287275	1

Table 35 – Macaw (Sad)

Macaw Sad	EMG	Humanness	Eeriness	Attractiveness
EMG	1			
Humanness	0.11847191	1		
Eeriness	0.02403438	-0.2167283	1	
Attractiveness	-0.0621024	**0.69938697**	-0.1128993	1

11.4 Emotional Expression Avatars – Animated

Table 36 – Curls (Happy)

Curls Happy	EMG	Humanness	Eeriness	Attractiveness
EMG	1			
Humanness	0.296623418	1		
Eeriness	0.322535289	0.217074229	1	
Attractiveness	0.211752921	**0.700682223**	0.135916675	1

Table 37 - Curls (Sad)

Curls Sad	EMG	Humanness	Eeriness	Attractiveness
EMG	1			
Humanness	-0.02570789	1		
Eeriness	0.408930675	0.415439056	1	
Attractiveness	-0.11091247	**0.682806592**	0.416802215	1

Table 38 - Macaw (Happy)

Macaw Happy	EMG	Humanness	Eeriness	Attractiveness
EMG	1			
Humanness	0.07073	1		
Eeriness	-0.1069	-0.1993	1	
Attractiveness	0.21016	0.45872	0.01495	1

Table 39 - Macaw (Sad)

Macaw Happy	EMG	Humanness	Eeriness	Attractiveness
EMG	1			
Humanness	-0.319384304	1		
Eeriness	0.318029962	0.0026197	1	
Attractiveness	-0.207543997	**0.644303286**	0.067260735	1

Table 40 - Jacqueline (Happy)

Jacqueline Happy	EMG	Humanness	Eeriness	Attractiveness
EMG	1			
Humanness	0.002022673	1		
Eeriness	-0.12685865	0.419753757	1	
Attractiveness	-0.2720404	0.220066207	0.082199	1

Table 41 - Jacqueline (Sad)

Jacqueline Sad	EMG	Humanness	Eeriness	Attractiveness
EMG	1			
Humanness	-0.13752006	1		
Eeriness	-0.306849676	0.0050294	1	
Attractiveness	0.041903094	0.342295122	-0.114840638	1