Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognize the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form and Deposit Licence.
Mona Lisa in the Uncanny Valley

Jingwen Mao

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy.

School of Psychology

The University of Auckland

2019
Abstract

Development in machine learning and AI has enabled the creation of virtual characters that are almost indistinguishable from human. Nevertheless, near-human characters tend to fall into the “uncanny valley”. The exploratory work in this thesis has explored whether face images generated by hyper-realistic simulation – the Auckland Face Simulator (the AFS) are comparable to photographic faces as stimuli in facial expression of emotion studies. Chapter Two investigated the psychophysiological “threshold” of facial expression recognition, especially subtle expressions. The findings demonstrated that happiness and anger could be differentiated at the 20% intensity level as early as the N170 time window of visual processing. Further, happiness required at least 40% of the maximum intensity to be distinguished from neutral in the ERP; anger appears to require even higher intensity. Chapter Three provided the first demonstration of the ERP differences between photo and the AFS faces. The amplitude differences between photo and simulated faces (Xyza and Leah generated by the AFS) commenced around the N170 time window and sustained till 300ms post-stimulus onset. More specifically, both N170 and P200 were larger in amplitude for photo faces than for simulated faces. The findings also showed the practicality of using ERPs to measure realism levels in hyper-realistic synthetic faces. Chapter Four and Five were extended analysis based on data from Chapter Three but targeting different questions. Chapter Four tested whether the uncanny valley effect can occur between different avatars. The ERP findings revealed a specific component (the POz positivity), and the P200 appeared to be sensitive to between-avatar differences. Chapter Five tested whether individual perception in the degree of realism of the avatars would influence the ERP responses. The results showed that individuals who perceive simulated faces as more real than photo faces tend to anthropomorphize and accept simulations. Taken together, the findings in the current thesis demonstrated that hyper-realistic synthetic faces, although they have achieved
photorealistic appearance, may not pass the facial Turing test just yet. The current work also lays the foundations of the psychophysiological approach in researching hyper-realistic avatars and designing avatars that pass the uncanny valley.
Acknowledgement

Firstly, I would like to thank my supervisors Associate Professor Paul Corballis and Associate Professor Mark Sagar, and my advisor Professor John Sollers, for all of their guidance and help during my PhD. Your generous support, encouragement, and guidance have smoothed my path towards my PhD. I would like to thank Paul for always being supportive to my interests and available to discuss any questions I had. The hyper-realistic avatars were provided by Mark and his team, and I would like to thank him for giving me the opportunity to work on and to test such a state-of-the-art face simulation – the Auckland Face Simulator. John provided me with help in designing my studies and also helped with statistical suggestions.

Particular appreciation goes to Lenore Tahara-Eckl, Dr. Matt Oxner, Nidhi Aggarwal, and Tasi Watson for providing valuable suggestions and feedback to my thesis. It is always a comfort to hear encouragement from them.

I am grateful to all my friends and family for their constant support and encouragement. In particular, I would like to give special mention to my sister Yiwen Mao for always being there for me, and my parents for always believing in me. Their love and guidance are with me in whatever I pursue.
# Table of Contents

Abstract ......................................................................................................................................................... i

Acknowledgement ......................................................................................................................................... iii

List of Figures .................................................................................................................................................. viii

List of Tables ................................................................................................................................................ xv

Chapter 1 General Introduction .................................................................................................................. 1

1.1 Defining emotion....................................................................................................................................... 3

1.2 Theories of Emotions and emotional expression ...................................................................................... 5

1.2.1 Discrete Models of Emotion / Basic Emotion Theories ........................................................................ 5

1.2.2 Dimensional Models of Emotion ........................................................................................................ 10

1.2.3 Convergence of the two models .......................................................................................................... 12

1.3 Emotion and facial expression ................................................................................................................ 14

1.4 Human Neuroscience of Emotion and Facial Expression ........................................................................ 16

1.4.1 Amygdala.............................................................................................................................................. 17

1.4.2 Prefrontal Cortices (PFC) .................................................................................................................. 18

1.4.3 Anterior Cingulate Cortex (ACC) and Insular Cortex ......................................................................... 19

1.4.4 Other Regions ..................................................................................................................................... 21

1.5 Left and Right Hemisphere Lateralization ............................................................................................... 23

1.6 ERP Components of Facial Expressions ................................................................................................ 25

1.7 FACS ...................................................................................................................................................... 26

1.8 Human-Computer Interaction and Human Avatar .................................................................................. 28

1.9 The Auckland Face Simulator ................................................................................................................ 29

1.9.1 Facial Expression and Motor System .................................................................................................. 30

1.9.2 Brain Language ................................................................................................................................ 32

1.9.3 BabyX – A Virtual Infant ................................................................................................................ 33

1.9.4 Xyza and Leah .................................................................................................................................. 36

1.10 Research Questions and Hypothesis ..................................................................................................... 40
Chapter 2 Identifying Subtle Expressions

2.1 Introduction

2.1.1 Facial Expression of Emotion Processing

2.1.2 Subtle vs Overt Facial Expressions

2.1.3 ERP and facial expression

2.1.4 The current study

2.2 Methodology

2.2.1 Participants

2.2.2 Materials

2.2.3 Procedure

2.2.4 EEG Recordings

2.3 Result

2.4 Discussion

2.4.1 Lateralization Effects on the P1

2.4.2 Emotional Effects on the N170

2.4.3 Facial Expression Processing over Frontal regions

2.4.4 Threshold of Identifying an Expression

2.4.5 Facial Expression Processing

2.4.5 Limitation and Future Direction

Chapter 3 The Uncanny Valley

3.1 The Emerging of Virtual Stimuli in Facial Expression Study

3.1.1 The Uncanny Valley in Robotic/Simulated Faces

3.1.2 Explanations of Uncanny Valley

3.1.3 Emotional Effects of The Uncanny Valley

3.1.4 Psychophysiological Studies of the Uncanny Valley

3.1.5 Efforts to Cross the Uncanny Valley

3.1.6 The Current Study
3.2 Method .................................................................................................................. 117
  3.2.1 Participants ....................................................................................................... 117
  3.2.2 Materials ......................................................................................................... 117
  3.2.3 Procedure ........................................................................................................ 117
3.3 Result...................................................................................................................... 120
  3.3.1 ERP Results ..................................................................................................... 120
  3.3.2 Self-report Questionnaire Result ...................................................................... 137
3.4 Discussion ............................................................................................................. 143
  3.4.1 Survey Findings ............................................................................................... 143
  3.4.2 ERP Findings ................................................................................................... 145
  3.4.3 Comparison between survey and ERP findings ................................................. 151
  3.4.4 The uncanny valley processing ........................................................................ 152
  3.4.5 Limitations ...................................................................................................... 153

Chapter 4 Model Differences ..................................................................................... 154
  4.1 Introduction ......................................................................................................... 154
  4.2 Methods .............................................................................................................. 156
4.3 Result..................................................................................................................... 156
  4.3.1 ERP Findings ................................................................................................... 161
  4.3.2 Self-report Questionnaire Result ...................................................................... 169
4.4 Discussion and Assumption .................................................................................. 173
  4.4.1 N170 and POz Positivity .................................................................................. 174
  4.4.2 P200 ............................................................................................................... 177
  4.4.3 Influences of Attractiveness and Age ................................................................ 178
  4.4.4 Superstimuli .................................................................................................... 181
  4.4.5 Next Steps ...................................................................................................... 184
  4.4.6 Conclusion ...................................................................................................... 184

Chapter 5 Group Differences ..................................................................................... 186
List of Figures

Figure 1.1 Mona Lisa .......................................................................................................................... 2
Figure 1.2 Xyza .................................................................................................................................37
Figure 1.3 Leah .................................................................................................................................38
Figure 2.1 The Bruce and Young Face Recognition Model (from Bruce and Young, 1986) .44
Figure 2.2 Trial procedure and the stimuli of the experiment .......................................................57
Figure 2.3 Grand average waves collapsed across emotions (happiness and anger) and
intensities (0%, 20% and 40%). Shaded areas represent ERP components of interest (in
temporal order - Fz: the FcEP Early, the FcEP Late I&II; P7 & P8: the P1 and the N170)....61
Figure 2.4 a) Scalp distributions for the P1 component (85 – 110ms) of the anger condition.
Topographic maps from left to right in order are 0%, 20% and 40% angry expression
respectively. b) Waveforms from left to right corresponding to P7 and P8 electrode clusters
chosen from posterior sites. The dashed rectangles indicate the magnified P1 component from
the waveform for better visualization. .........................................................................................63
Figure 2.5 a) Scalp distributions for the P1 component (85 – 110ms) of the happiness
condition. Topographic maps from left to right in order are 0%, 20% and 40% happy
expression respectively. b) Waveforms from left to right corresponding to P7 and P8
electrode clusters chosen from posterior sites. The dashed rectangles indicate the magnified
P1 component from the waveform for better visualization. .......................................................64
Figure 2.6 Bars show the amplitude of the P1 component for each expression condition at left
(P7 cluster) and right (P8 cluster) scalp locations. The curved bracket illustrates the
significant main effect between left and right electrode clusters. *p < 0.05. Error bars show
SEM. ...........................................................................................................................................65
Figure 2.7 a) Scalp distributions for the N170 component (120 – 160ms) of the anger
condition. Topographic maps from left to right in order are 0%, 20% and 40% angry
expression respectively. b) Waveforms from left to right corresponding to P7 and P8
electrode clusters chosen from posterior sites. The dashed rectangles indicate the magnified
N170 component from the waveform for better visualization....................................................66
Figure 2.8 a) Scalp distributions for the N170 component (120 – 160ms) of the happy
condition. Topographic maps from left to right in order are 0%, 20% and 40% happy
expression respectively. b) Waveforms from left to right corresponding to P7 and P8 electrode clusters chosen from posterior sites. The dashed rectangles indicate the magnified N170 component from the waveform for better visualization.................................67

Figure 2.9 Bars show the amplitude of the N170 component for each expression condition at left (P7 cluster) and right (P8 cluster) scalp locations. The curved bracket illustrates the significant main effect between anger and happiness emotions. *p < 0.05. Error bars show SEM. ..................................................................................................................68

Figure 2.10 a) Scalp distributions for the FcEP early component (120 – 160ms) of anger and happiness condition. Topographic maps from left to right in order are 0%, 20% and 40% for angry and happy expression respectively. b) Waveforms from left to right corresponding to Fz electrode clusters chosen from anterior sites for anger and happy condition. The dashed rectangles indicate the magnified FcEP early component from the waveform for better visualization. c) Bars show the amplitude of the FcEP early component for each expression. The curved bracket illustrates the significant main effect between anger and happiness emotions. *p < 0.05. Error bars show SEM.................................................................69

Figure 2.11 a) Scalp distributions for the FcEP late I component (210 – 295ms) of anger and happiness condition. Topographic maps from left to right in order are 0%, 20% and 40% for angry and happy expression respectively. b) Waveforms from left to right corresponding to Fz electrode clusters chosen from anterior sites for angry and happy condition. The dashed rectangles indicate the magnified FcEP late I component from the waveform for better visualization. c) Bars show the amplitude of the FcEP late I component for each expression. Error bars show SEM.................................................................71

Figure 2.12 a) Scalp distributions for the FcEP late II component (295 – 380ms) of anger and happiness condition. Topographic maps from left to right in order are 0%, 20% and 40% for angry and happy expression respectively. b) Waveforms from left to right corresponding to Fz electrode clusters chosen from anterior sites for angry and happy condition. c) Bars show the amplitude of the FcEP late II component for each expression. Error bars show SEM..............72

Figure 3.1 Examples of Ohman et al.'s (2001) schematic faces........................................88

Figure 3.2 Examples of face models generated by FaceGen with different skin texture. .......90

Figure 3.3 Illustration of Mori’s uncanny valley curve showing a non-linear relationship between human likeness and affinity. Mori hypothesized that movement will amplify the uncanny valley. Adapted from Mori (1970) translated by MacDorman and Kageki (2012). .93
Figure 3.4 Demonstration of examples of the two sides of the uncanny valley: the left-hand machine side and the right-hand human side.

Figure 3.5 AI-generated images of celebrities.

Figure 3.6 Experimental procedure of the study and the stimuli of the experiment.

Figure 3.7 Grand average waves of photo and simulation conditions. Shaded areas represent ERP components of interest (P7 and P8 clusters: the P1 and the N170; POz: the POz Positivity; O1, Oz and O2: the P200).

Figure 3.8 a) Individual scalp distributions for P1 component (85-110ms) of all conditions: from top to bottom are 0% (neutral), 20% and 40% facial expressions for both photo and simulation conditions. Topographic maps on the left depict photo conditions and on the right show simulation conditions respectively. b) Waveforms from left and right portray grand averages of photo and simulation conditions at the P7 (left) and P8 (right) electrode clusters from posterior sites. The shaded area on the waveforms reflects the P1 component. c) Bars display the amplitude of the P1 for anger and happiness with each intensity for photo and simulated conditions at left (P7 cluster) and right (P8 cluster) scalp locations. Error bars show SEM.

Figure 3.9 a) Individual scalp distributions for N170 component (120-160ms) of all conditions: from top to bottom are 0% (neutral), 20% and 40% facial expressions for both photo and simulation conditions. Topographic maps on the left depict photo conditions and on the right show simulation conditions respectively. b) Waveforms from left and right portray grand averages of photo and simulation conditions at the P7 (left) and P8 (right) electrode clusters from posterior sites. The shaded area on the waveforms reflects the N170 component. c) Bars display the amplitude of the N170 for anger and happiness with each intensity for photo and simulated conditions at left (P7 cluster) and right (P8 cluster) scalp locations. The curved brackets indicate significant main effects of electrode and realism. *p < 0.05. Error bars show SEM.

Figure 3.10 a) Individual scalp distributions for P200 component (170-280ms) of all conditions: from top to bottom are 0% (neutral), 20% and 40% facial expressions for both photo and simulation conditions. Topographic maps on the left depict photo conditions and on the right show simulation conditions respectively. b) Waveforms from left and right portray grand averages of photo and simulation conditions at the O1 (left), Oz (middle) and O2 (right) electrode from occipital sites. The shaded area on the waveforms reflects the P200
component. c) Bars display the amplitude of the P200 for anger and happiness with each intensity for photo and simulated conditions at the O1 (left), Oz (middle) and O2 (right) electrode scalp locations. The curved brackets indicate significant main effects of electrode, realism and emotion. *p < 0.05. Error bars show SEM. ................................................................. 130

Figure 3.11 a) Individual scalp distributions and waveforms for the FcEP late I (210-295ms) of all conditions: from top to bottom are 0% (neutral), 20% anger, 40% anger, 20% happiness, and 40% happiness for both photo and simulation conditions. Topographic maps on the left depict photo conditions and on the right show simulation conditions respectively. b) The waveform portrays grand averages of photo and simulation conditions at the Fz electrode clusters from anterior sites. The shaded area on the waveform reflects the FcEP late I. c) Bars display the amplitude of the FcEP late I for anger and happiness with each intensity for photo and simulated conditions. The curved brackets indicate significant main effects of emotions and realism. *p < 0.05. Error bars show SEM. ................................................................. 133

Figure 3.12 a) Individual scalp distributions and waveforms for the FcEP late II (295-380ms) of all conditions: from top to bottom are 0% (neutral), 20% anger, 40% anger, 20% happiness, and 40% happiness for both photo and simulation conditions. Topographic maps on the left depict photo conditions and on the right show simulation conditions respectively. b) The waveform portrays grand averages of photo and simulation conditions at the Fz electrode clusters from anterior sites. The shaded area on the waveform reflects the FcEP late II. c) Bars display the amplitude of the FcEP late II for anger and happiness with each intensity for photo and simulated conditions. Error bars show SEM. ................................................................. 136

Figure 3.13 a) Scatter plot of photo and simulation indicates human likeness score. b) Bars display the human likeness score of anger and happiness with 3 levels of intensity (0%, 20% and 40%) for photo and simulation models. The curved brackets indicate a significant main effect of intensity. *p < 0.05. Error bars show SEM. ................................................................. 139

Figure 3.14 Bars display the correct response rate of anger and happiness with 3 levels of intensity (0%, 20% and 40%) for photo and simulation models. The curved brackets indicate significant main effects of model, emotion and intensity. *p < 0.05. Error bars show SEM 142

Figure 4.1 ERP waveforms collapsed across all emotional expressions and intensities compared between Leah, Xyza and Photo. The waveforms are corresponding to the electrodes (clusters) presented in Table 4.1. ................................................................. 158

Figure 4.2 A bird’s eye view of the scalp map with electrodes of interests. ............................. 159
Figure 4.3 a) Scalp distributions for the N170 and POz Positivity components (120-160ms) of the grand average: from left to right are Leah, Xyza and Photo models. b) Waveforms from left to right represents grand averages of Leah, Xyza and Photo at the P7 (left), POz (middle) and P8 (right) clusters from posterior sites. The shade area on the waveforms reflects the N170 component (at P7 and P8) and POz positivity (at POz). c) Bars display the amplitude of the N170 and the POz Positivity for Leah, Xyza and photo. The curved brackets indicate significant main effects of model. *p < 0.05. Error bars show SEM. .......................................................... 163

Figure 4.4 a) Scalp distributions for the P200 (170-280ms) for the grand averages of three models: from left to right are Leah, Xyza and Photo. b) Waveforms from left to right represent grand averages of Leah, Xyza and photo at the O1 (left), Oz (middle), and O2 (right) clusters from occipital sites. The shaded area on the waveforms reflects the P200 component. c) Bars display the amplitude of the P200 for Leah, Xyza and photo over occipital sites: O1 (left), Oz (middle), and O2 (right). The curved brackets indicate significant main effects of site and model. *p < 0.05. Error bars show SEM. ................................. 166

Figure 4.5 Bars represented the human-likeness score of Leah, Xyza and Photo for anger and happiness with three intensity levels (0%, 20%, and 40%). The curved brackets indicate significant main effect of intensity. *p < 0.05. Error bars show SEM. ........................................... 169

Figure 4.6 Bars represent the correct response rate of Leah, Xyza and Photo for anger and happiness with three intensity levels (0%, 20% and 40%). The curved brackets indicate significant main effects of model, emotion and intensity. *p < 0.05. Error bars show .......... 172

Figure 4.7 Schematic baby faces in McKelvie’s (1993) study. Face 1 has larger eyes and low feature placement. Face 2 has the same feature placement with Face 1 except for smaller eyes. Face 3 moves features in Face 1 upwards. Face 4 has the same feature placement with Face 3 except for smaller eyes. ......................................................................................... 183

Figure 5.1 a) Scalp distributions for the N170 (120-160ms) for the grand averages of three groups: from left to right are SLR, SEP and SMR. b) Waveforms from left to right represent grand averages of the three groups (SLR, SEP and SMR) at the P7 (left) and P8 (right) clusters from posterior sites. The shaded area on the waveforms reflects the N170 component. c) Bars display the amplitude of the N170 for groups (SLR, SEP and SMR), and realism (photo and simulation). The curved brackets indicate significant main effects of group and realism. *p < 0.05. Error bars show SEM. .......................................................................................... 192
Figure 5.2 a) Scalp distributions for the P200 (170-280ms) for the grand averages of three groups: from left to right are SLR, SEP and SMR. B) Waveforms from left to right represent grand averages of the three groups (SLR, SEP and SMR) at the O1 (left), Oz (middle) and O2 (right) electrodes from occipital sites. The shaded area on the waveforms reflects the P200 component. c) Bars display the amplitude of the P200 for groups (SLR, SEP and SMR), and realism (photo and simulation). The curved brackets indicate a significant main effect of realism. *p < 0.05. Error bars show SEM.

Figure 5.3 a) Scalp distributions for the N170 (120-160ms) for the grand averages of three groups: from left to right are LH, MH and HH. B) Waveforms from left to right represent grand averages of the three groups (LH, MH and HH) at the P7 (left) and P8 (right) electrode clusters from posterior sites. The shaded area on the waveforms reflects the N170 component. c) Bars display the amplitude of the N170 for groups (LH, MH and HH), and realism (photo and simulation). The curved brackets indicate a significant main effect of realism. *p < 0.05. Error bars show SEM.

Figure 5.4 a) Scalp distributions for the P200 (170-280ms) for the grand averages of three groups: from left to right are LH, MH and HH. B) Waveforms from left to right represent grand averages of the three groups (LH, MH and HH) at the O1 (left), Oz (middle) and O2 (right) electrodes from occipital sites. The shaded area on the waveforms reflects the P200 component. c) Bars display the amplitude of the P200 for electrodes (O1, Oz and O2), groups (LH, MH and HH), and realism (photo and simulation). The curved brackets indicate significant main effects of electrode and realism. *p < 0.05. Error bars show SEM.

Figure 5.5 a) Scalp distributions for the N170 (120-160ms) for the grand averages of three groups: from left to right are LS, MS and HS. B) Waveforms from left to right represent grand averages of the three groups (LS, MS and HS) at the P7 (left) and P8 (right) electrode clusters from posterior sites. The shaded area on the waveforms reflects the N170 component. c) Bars display the amplitude of the N170 for groups (LS, MS and HS), electrodes (P7 and P8 clusters) and realism (photo and simulation). The curved brackets indicate significant main effects of electrode and realism. *p < 0.05. Error bars show SEM.

Figure 5.6 a) Scalp distributions for the P200 (170-280ms) for the grand averages of three groups: from left to right are LS, MS and HS. B) Waveforms from left to right represent grand averages of the three groups (LS, MS and HS) at the O1 (left), Oz (middle) and O2 (right) electrodes from occipital sites. The shaded area on the waveforms reflects the P200
component. c) Bars display the amplitude of the P200 for groups (LS, MS and HS), and realism (photo and simulation). The curved brackets indicate a significant main effect of realism. *p < 0.05. Error bars show SEM.
List of Tables

Table 1.1  Discrete emotions and corresponding neural pathways ......................................................8
Table 2.1  Details of Components Analyzed .........................................................................................60
Table 3.1  Details of Components Analyzed .........................................................................................121
Table 4.1  Electrodes (clusters) of interests .........................................................................................159
Table 4.2  N170 and POz Positivity ....................................................................................................160
Chapter 1 General Introduction

Mona Lisa (see Figure 1.1) is probably one of the most ambiguous and controversial art pieces in the world. The debates may root from her mysterious smile, or to be specific, is she smiling? From some viewpoints people see Mona Lisa smiling, yet from some other angles, the smile disappears. More often, you will see her smiling but when you avert your gaze, she seems to portray a sad stare (Squires, 2015). What contributes to her enigmatic smile? Would her smile be uncatchable if she had her mouth widely open and cheek raised? The answer seems to be no. Therefore, presumably, the mystery probably originates in the subtlety of Mona Lisa’s expression. Kornmeier and Mayer (2014) argued that such ambiguity is derived from the limited capacity of human’s senses to handle ambiguous perceptual information, and the brain has to construct the information and interpret in a meaningful way. When faced with insufficient or ambiguous sensory information, Mona Lisa’s facial expression becomes changeable and enigmatic.

Besides the emotional ambiguity, Mona Lisa’s following stare towards viewers is often described as “creepy”. She was no doubt a real person, but the interesting features of the painting produce an unreal feeling of her, which can make viewers uneasy. This has fascinated us to think about how do we as humans perceive realistic but simultaneously unreal agents.

Mona Lisa further makes researchers wonder whether subtle depictions of emotion influence the perception of humanness in ambiguous characters. In the following sessions of this chapter, I will explore theories and neuroscience underneath emotion and facial expression. Furthermore, I will also encompass concepts of human avatar, Face Action Coding System (FACS) and introduce the Auckland Face Simulator (AFS).
Figure 1.1 Mona Lisa

Source: LeonardoDaVinci.net (2019)
1.1 Defining emotion

In daily life, the term emotion usually means feelings, such as happy, angry, love, or contempt. However, scientifically, it is difficult to define emotion since it is a vague and variable concept. Researchers tend to investigate different aspects based on the way they interpreted what an emotion is. Some researchers have looked at emotion as a feeling or reaction where they ask “how sad, happy, fear, disgust, anger or surprise do you feel now on a scale from 1 to 10” (Scherer, 1984). They usually administered self-report questionnaires and presumed that participants were able to qualify and reflect these emotional states accurately, which in turn implied that they were conscious of the emotional states. Other researchers have interpreted emotion in terms of facial expressions and looked into the methodology to recognize, interpret, and measure those facial expressions (Ekman, Hager, & Friesen, 1981). In addition, another group of researchers defines emotion physiologically based on physiological results and responses from nervous system such as heart rate, the SCR (skin conductance responses) or electrodermal responses (Appelhans & Luecken, 2006; Esteves, Dimberg, & Ohman, 1994; Gendolla, Abele, & Krusken, 2001). Both physiological and expressional aspects of emotion suggested that emotions could occur unconsciously. Purves, et al. (2013, p. 321) provided a relatively thorough definition of emotions, which are “sets of physiological responses, action responses, and subjective feelings that adaptively engage humans and other animals to react to events of biological and/or individual significance”.

Another reason why emotion is difficult to define lies in different beliefs on how an emotion occurred. Started with Darwin (1872/1998; Cosmides & Tooby, 2000) who proposed an evolutionary prospective to explain the occurrence of emotions, researchers within this paradigm argued that the emotional responses of human are biologically prepared during evolution. Nesse (1990) noted that emotions were shaped by natural selection. When people
cannot display normal emotions or their emotions malfunction, they will be disrupted in their social and everyday life. Other researchers embraced a more functional approach of emotion occurrence where they were convinced that emotions are induced within a certain environment with specific events ongoing and the necessity in communicating with other individuals (Keltner & Gross, 1999; Niedenthal & Brauer, 2012). More specifically, Keltner and Gross (1999) conceptualized the origin of emotion on functional accounts where: (1) emotion served as a solution to survival-relevant problems, for instance, to maintain a relationship or avoid intimidation (Plutchik, 1984; Scherer, 1994); (2) emotion acts as a combination of several systematic coordinated components. Emotional behavioral cues contribute to purposeful communication, emotion perception and experience, serve to influence information processing and decision making. Emotional regulation facilitates adaptation to the social and physical environments (Hareli & Parkinson, 2008; Keltner & Haidt, 1999). (3) The functions of emotion are often parallel to beneficial consequences where scholars studied emotional experience, which improved surviving of individuals and enhanced both physical and social environment (Carstensen, Gross, & Fung, 1997; Zech & Rime, 2005). As a result, Keltner and Gross (1999, p. 468) gave out a rather broad definition of emotion, which is “episodic, relatively short-term, biologically-based patterns of perception, experience, physiology, action, and communication that occur in response to specific physical and social challenges and opportunities.”
1.2 Theories of Emotions and emotional expression

Understanding the fundamental mechanism of emotions and facial expressions is always an ongoing task in the affective field. Despite the efforts of extensive studies in the topic through various methods, it remains a debate of whether emotions are constructed from distinct entities or continuous dimension properties, which forms the two dominant theories, namely discrete/basic emotion theory and dimensional theory.

1.2.1 Discrete Models of Emotion / Basic Emotion Theories

Discrete model argued that emotions, especially basic emotions, are biologically distinctive, and even universal across species (Purves et al., 2013). The prominent foundation of the discrete model derives from Darwin (1872/1998), where during evolution some basic or fundamental emotions have developed their own specific “eliciting conditions and its own specific physiological, expressive, and behavioral reaction patterns” (Scherer, 2000, p. 147). In other words, emotions can be categorized into several basic ones, including happiness, anger, fear, surprise, disgust, and sadness (Ekman, 2003), which also pinpoint distinct expressive pathways of emotions.

The fundamental assumption underpinned is that every basic emotion corresponds to a discrete neural system. More generally, basic emotions are thought to result in different cognitive functions, from sensory and perceptual cognitive processes to physiological changes (Scherer, 2000b). To illustrate, a number of lines of evidence strongly associate the insula and the basal ganglia to disgust, the medial prefrontal cortices to happiness, sadness and fear to the amygdala (Murphy, Nimmo-Smith, & Lawrence, 2003) and anger to the orbitofrontal cortex (OFC) (Blair & Cipolotti, 2000). Likewise, Pitcher (2014) used transcranial magnetic simulation (TMS) to disrupt the right posterior superior temporal sulcus (rpSTS) and the right occipital face area (rOFA) in a facial expression recognition task. They
found that the impairment induced by the TMS between the rpSTS and rOFA has latency differences, which both rpSTS and rOFA impaired task performance at 60-100ms window while the impairment effect of rpSTS sustained 40ms more. Pitcher (2014) interpreted such latency difference as functionally distinct neural computations of facial expression recognition in each region.

Tettamanti et al. (2012) found that a distributed set of brain regions have responded to all ranges of emotional stimuli compared to neutral, which include “the right amygdala, associative occipital and temporal regions, and the cerebellum” (p. 1811). They tested four basic emotions, namely happiness, disgust, sadness, and fear. The fMRI results were able to identify distinct neural arrays to single basic emotions, with the exception of sadness. Besides, Tettamanti and colleagues (2012) have employed dynamic causal models (DCM) to examine how common brain regions and domain-specific regions overlap or interact with each other. To analyze the functional specialization, they only considered the brain region that is exclusively activated in a single basic emotion using a stringent approach. Except sadness, they were able to identify domain-specific regions corresponding to three emotions: the frontoparietal system is thought to prepare motor responses in fear conditions; the somato-sensory system activates protective responses in disgust situations; and the medial prefrontal and temporoparietal cortices respond to happy interactions.

Kragel and LaBar’s (2015) study using fMRI has supported that emotions are organized and categorized in the brain with functionally distinct pathways (see
They highlighted that emotions with similar dimensional index, such as contentment and amusement, which were both associated with positive and pleasant state, were not sharing the same region of the brain. Activation in precuneus, medial prefrontal, cingulate and primary somato-sensory cortices was mapped on contentment, whereas temporal cortex, supplementary motor area, and thalamus were activated in processing amusement (Northoff et al., 2006; Wild et al., 2006). Apart from that, Kragel and LaBar (2015) employed the multivariate statistical method and Bayesian approach to computational model and predicted emotional states on neural patterns. They found that those emotional states that differ in dimensions were difficult to predict mathematically, while differences in discrete emotional states displayed better classification in terms of neural responses. However, they also furthered that valence and arousal aid in explaining confusion of emotions and discriminating among emotions.
<table>
<thead>
<tr>
<th>Emotions</th>
<th>Activated Brain Regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contentment</td>
<td>Dorsal precuneus, bilateral postcentral gyrus and mid-cingulate gyrus</td>
</tr>
<tr>
<td>Amusement</td>
<td>Bilateral superior- and middle-temporal lobe, bilateral early visual cortex and bilateral supplementary motor area, among other frontal regions</td>
</tr>
<tr>
<td>Surprise</td>
<td>Mid- and anterior-cingulate gyrus, bilateral anterior insula, bilateral mid-occipital cortex, bilateral thalamus and bilateral cerebellum</td>
</tr>
<tr>
<td>Fear</td>
<td>Bilateral lingual gyrus, bilateral fusiform gyrus, dorsal/anterior precuneus and a number of bilateral medial temporal lobe structures: amygdala, hippocampus and parahippocampal cortex</td>
</tr>
<tr>
<td>Anger</td>
<td>Bilateral superior temporal gyrus, ventral precuneus and right angular gyrus</td>
</tr>
<tr>
<td>Sadness</td>
<td>Left cerebellum, bilateral superior temporal gyrus and bilateral temporal pole</td>
</tr>
<tr>
<td>Neutral</td>
<td>Bilateral activation in angular gyrus, supramarginal gyrus, postcentral gyrus and lingual gyrus</td>
</tr>
</tbody>
</table>

*Source: Kragel and LaBar (2015, pp. 1441-1443)*
Saarimaki et al. (2016) employed multivariate pattern analysis (MVPA) with fMRI using affective videos or emotional imagery and found distinctive but broad neural distribution of all six basic emotions. These neural signatures were sustained in both internal and external emotional stimulation conditions, but the generalization of the neural activity pattern leans towards biological rather than experience-based emotions. However, they were unable to identify a one-to-one correspondence between an emotion and a brain region. They thus concluded that the organization of emotions in the brain is thus characterized by “spatially overlapping but distinct local activation patterns” (Saarimaki et al., 2016, p. 2564).

Costa et al.’s (2014) study on time course may also give some insights on how different emotions were constructed. They conducted an ERP study to investigate the temporal signatures of discrete emotions. They had participants passively view images from the IAPS (the International Affective Picture System). Participants were required to categorize the images into four emotions (fear, disgust, happiness, and sadness) and rate the valence and arousal. The findings showed that fear occurs at the earliest time segments, followed by disgust and happiness, and sadness appears to be the slowest. Moreover, the localization analysis also found two time segments for significant activations of fear, disgust, and happiness, while only one but sustained time segment was found in sadness. Likewise, Grootswagers, Kennedy, Most and Carlson (2017) employed MEG to study the temporal characteristics of different emotions. To avoid the stimuli being predefined categorically, they have participants rated emotional images on valence and arousal, and intensity of anger, sadness, fear, disgust and happiness as well. They applied representational similarity analysis (RSA) to examine the time course of emotional responses. Their findings suggested that the brain reacts to both arousal and valence fairly early regardless of various arousal and valence levels in the stimuli. Responses to fear, disgust and happiness appeared earlier than anger, and the response to sadness was not reliably captured during their time window. In addition,
they reported that different emotions are differentiable by RSA in early time course. When adjusting the valence and arousal ratings, distinct fear and disgust representations were still present. Thus, the findings leaned towards discrete emotional constructs.

1.2.2 Dimensional Models of Emotion

In contrast to discrete theory of emotion, dimensional model argued that each emotion corresponds to a point on a coordinate that is pinpointed by two axes, such as arousal and valence (Russell, 1989) or approach and withdrawal (Davidson & Irwin, 1999). Russell (1989) assumed that emotions are perceived as subjective experience on a continuum along two dimensions, namely valence and arousal. Valence is a subjective feeling of positive or negative feelings while arousal is to measure if the subjective feeling is activated or deactivated. The emotional experience can be portrayed by where these emotions positioned along the axis. Sadness, for instance, is characterized by negative valence and low arousal, while anger or fear is characterized by negative valence and high arousal (Russell, 1989).

Another dual-dimensional model was proposed by Davidson and Irwin (1999), which is the approach-withdrawal model. The approach dimension represents goal-directed behaviors and generates approach-related positive emotions (e.g., pride), whereas the withdrawal dimension facilitates withdrawal behaviors from aversive stimulation and generates withdrawal-related negative emotions (e.g., fear and disgust).

Posner, Russell and Peterson (2005) proposed the circumplex model of affect, which assumes that emotional states are rooted in two neurophysiological systems that correspond to valence (unpleasant-pleasant) and arousal (high-low). They positioned each emotion with a coordinate of valence and arousal. Happiness is characterized by high arousal and high pleasant, and sadness locates at a combination of low arousal and low unpleasant. For unpleasant emotions, fear has the highest arousal, followed by anger and then sadness.
A series of studies involving patients with brain damage, emotional disorder, other clinical populations and even rhesus monkeys (Davidson, 1984; Heller, Nitschke, & Miller, 1998; Kalin, Larson, Shelton, & Davidson, 1998; Robinson & Manes, 2000) have postulated that positive and negative emotions are regulated by at least partially distinctive neural substrates. Studies reported that patients with right-hemisphere or anterior cerebral lesions have a tendency to be euphoric (Gainotti, 1972). On the contrary, patients with left-hemisphere lesions tend to enhance negative facial expression production (Sackeim et al., 1982).

Although in discrete theories, the mapping of each basic emotion on specific brain regions is under debate, and the association between emotional valance and arousal and brain regions are also elusive. More recently, emerging neuroimaging studies have proposed the separation of neural systems in processing valence and arousal. The presence of unpleasant pictures has been associated with activation in the amygdala, the right hippocampus and the medial occipital lobe (Gerdes et al., 2010; Weierich, Wright, Negreira, Dickerson, & Barrett, 2010), visual and lateral prefrontal regions (Nielen et al., 2009), while pleasant pictures activated left occipital regions and partial medial temporal lobe (Gerdes et al., 2010), and orbitofrontal regions (Nielen et al., 2009). Research further demonstrated a distinct neural network for valance and arousal of emotions, where midline and medial temporal lobe subserve arousal and dorsal cortical and mesolimbic regions mediate valence (Colibazzi et al., 2010). Further, the amygdala (Gerdes et al., 2010) and the anterior cingulate cortex (Viinikainen et al., 2010) have been shown to activate more strongly with relatively high-arousing unpleasant pictures, whereas the caudate, the nucleus accumbens (Gerdes et al., 2010) and the right substantia innominate (Viinikainen et al., 2010) were activated stronger with higher arousing pleasant pictures.
1.2.3 Convergence of the two models

In general, emotions are far more complex than just the basic emotions. This is characterized by the fact that facial features distinguishing between some emotional expressions are usually subtle. For instance, individuals typically do not perform well in differentiating between disgust and contempt or fear and surprise (Ekman & Friesen, 1971). Furthermore, existing studies have not shown a clear and consolidated link between all basic emotions and specific brain regions. The neural regions or systems that tie specifically to sadness or surprise (Murphy et al., 2003; Tettamanti et al., 2012) remain unclear. On the other hand, it is probably not inequitable to think that emotions may be discrete in the sense of different activation pattern or neural routes instead of regions (Tettamanti et al., 2012). Similarly, Costa et al. (2014) also challenged the idea of one-to-one mapping between a single basic emotion and specific brain regions. They emphasized that brain regions do not function in an isolated manner. Rather, multiple regions can activate depending on which network they belong to. Further, the functional specialization of a neural subtract does not necessarily differentiate spatially, it can also be differentiated temporally. Taken together, multiple neural systems co-activate when perceiving a specific emotion in distinct time windows could also suffice the concept of discrete emotion theories.

This idea could be further supported by findings suggesting the midline regions (including medial prefrontal, medial posterior regions, precuneus, and posterior cingulate cortex) (Saarimaki et al., 2016) and limbic regions are activated in multiple emotions. These regions may act like a “central executive” headquarter for emotional processing. When the input of different emotional signals is received and pre-processed from different pathways, the “central executive” will further process these signals and send feedback accordingly via different neural substrates, which then constitutes discrete emotions. Moreover, neural substrates may overlap with similar expressions (e.g., fear vs. sadness, positive or negative
emotions). An analogy to this could be a highway/motorway, where the target locations (analogous to emotions) are discrete, the routes overlap but are also separated.

Furthermore, there are also some criticisms regarding the dimensional approach of emotions. Firstly, both valence and arousal are measured by ratings, where valence ratings are more consistent than arousal ratings (Calvo & Nummenmaa, 2016). Scholars agree that happiness is a pleasant emotion while anger, sadness, fear, and disgust are more unpleasant than neutral emotions. Arousal ratings, on the other hand, produced more inconsistent results. Happiness was found to be equally arousing with anger and fear, whereas surprise is less arousing (Adolph & Alpers, 2010). In another ERP study, happiness was reported to be more arousing than fear and sadness, but equally arousing with anger (Calvo & Beltran, 2013). Secondly, dimensional models are unable to position some basic emotions. Emotions like happiness, anger, or fear are well located along two dimensions, but surprise or contempt, for example, seems difficult to locate in the dimension plane. Moreover, some emotions may be difficult to distinguish along the axes, such as anger and fear (Gunes & Pantic, 2010). Thirdly, emotional position alongside two dimensions can be ambiguous. Emotions such as anger, disgust, fear, and sadness are traditionally perceived as negative ones. Yet, individuals could differ in the level and nature of negativity in terms of their personal evaluation. To be more specific, one may perceive disgust as more negative than another, or, anger is more negative than sadness (Harmon-Jones, Harmon-Jones, Amodio, & Gable, 2011). This may then result in individual varieties in perceiving emotional valence. Hence, discrete emotions may appear to be more suitable and describable in coordinating individual differences.

Indeed, neither two models can independently account for emotional facial expression processing. Although the two emotional models seemingly oppose each other, emerging lines of studies showed that the two models may co-occur during facial expression processing but can be region-specific where the right fusiform face area (FFA) may respond to categorical
emotions, and the amygdala, insula and medial prefrontal cortex may mediate dimensional processing (Fujimura, Matsuda, Katahira, Okada, & Okanoya, 2012; Matsuda et al., 2013).

1.3 Emotion and facial expression

It appears that the concepts of emotion and facial expression are interchangeable and integrable. Indicatively, facial expression is a more explicit way to know how a person is feeling. However, the interaction between emotion and facial expression remains unclear. A good example may be to fabricate facial expressions with the absence of emotion. Individuals generally are unable to distinguish between false and genuine expressions (Ekman, 2003).

Ekman, Levenson and Friesen (1983) discovered that posing facial expressions could result in physiological elicitation of corresponding emotions, which constitutes the Facial Feedback Hypothesis. This hypothesis was later supported by Strack, Martin and Stepper (1988) who showed self-generated facial expressions change subjective emotional experience and Soussignan (2002) who reported consistent results. Further, several lines of evidence postulated inhibition of facial expression decreased subjective emotional experience and behavior (Davis, Senghas, & Ochsner, 2009; Goldin, McRae, Ramel, & Gross, 2008). Neal and Chartrand (2011) also found significant impairment of emotion perception in those who had plastic surgery (e.g., injecting Botox) that inhibits facial muscle moments. In principle, smiling (movement with the zygomatic major muscles) tends to elicit more positive emotions, while frowning (movement with corrugator supercilii muscle) induces negative emotions (Dimberg & Soderkvist, 2011).

Facial electromyography (EMG) studies provide a more direct relationship between emotion and facial expressions. Jancke (1996) recorded facial muscle activities while having participants experiencing anger by providing negative feedback on an intelligence test. Facial
EMG detected activities over the frontalis and corrugator muscles in those participants who verbally expressed anger. Another study (Lee et al., 2013) measured facial EMG of the corrugator muscle, the skin conductance response (SCR) and pupil size (PS) of participants who either imitate or observe angry facial expression, and observed greater responses to all three physiological measurements in imitation compared to observation of angry expression. These findings are evident in the communication between facial expressions and emotions.

Neuroimaging studies demonstrated that observing facial expressions can elicit both behavioral and neurological responses of emotional empathy. Perceiving facial expression of pain activates common neural network with subjective experience of pain (Benuzzi et al., 2018; Botvinick et al., 2005; Corradi-Dell’acqua, Hofstetter, & Vuilleumier, 2011). Wicker et al. (2003) found that visual process of the disgust expression evoked the internal state of disgust, and both activated the same brain areas, the insular in particular. On the other hand, Hennenlotter et al. (2008) applied botulinum toxin (BTX) to minimize the afferent muscular input. Participants were required to imitate angry facial expression, and the results yielded that BTX injection attenuates brain areas, including the left amygdala and brain stem regions. The study provides further physiological basis for conveying emotions via facial expressions.

Taken together, both theories and empirical studies provide some insights and evidence on the reflection and representation of emotions in terms of facial expressions.
1.4 Human Neuroscience of Emotion and Facial Expression

The seventh cranial nerve, also known as the facial nerve, controls the facial muscles of expressions. The nerve is regulated by two areas in the brain, which also correspond to voluntary and involuntary control of expressions. The pyramidal tract roots from the cortical motor strip and connects to facial muscles of voluntary facial actions, while the extrapyramidal tract roots from the subcortical areas and connects to facial muscles of involuntary facial actions (Cattaneo & Pavesi, 2013; Hwang & Matsumoto, 2016). The cortical motor strip is unique to humans, whereas the subcortical areas have evolved long ago in both humans and animals to deal with survival-related functions including feeding, reproduction, fighting and fleeing (Cattaneo & Pavesi, 2013; Hwang & Matsumoto, 2016).

Certain brain areas are demonstrated to be involved in the ability to recognize facial expressions via functional imaging and lesion studies. Those brain areas are largely distributed, including multiple networks of the occipito-temporal cortex (Adolphs, Damasio, Tranel, Cooper, & Damasio, 2000), the orbito-frontal cortex (Gorno-Tempini et al., 2001), the amygdala (Adolphs, Tranel, Damasio, & Damasio, 1994; Gorno-Tempini et al., 2001), the basal ganglia (Batty & Taylor, 2003) and the insula (Phillips et al., 1998). While some discrete emotions can pinpoint specific neural regions, studies have postulated that the rostral supracallosal anterior cingulate (ACC) (MacLean, 1993) and dorsal medial prefrontal cortices (PFC) (Damasio, 1994) were active in a range of emotions.

A more systematic line of studies focused on how the amygdala, the PFC, the ACC and insular cortex associate with emotional processing. I will also review studies on the relationship between other regions of the brain and emotions.
1.4.1 Amygdala

The amygdala is of anatomical and functional significance in emotion and facial expression processing. It is often known as a key subcortical structure of fear emotion and response (Armony, 2013; LeDoux, 2003). The amygdala-fear association was initially established from animal studies in Pavlovian fear conditioning and extinction (Li et al., 2013; Maren, Phan, & Liberonz, 2013). In humans, studies on amygdala lesions have revealed some functional roles of the amygdala. Patients with focal bilateral amygdala lesions showed impairments in reporting fear experience even while in contact with fearful stimuli (spiders and snakes) (Feinstein, Adolphs, Damasio, & Tranel, 2011).

Although the vast majority of amygdala studies centered on fear, researchers are growingly interested in the amygdala’s responses in other expressions. The study of patient SM who has focal bilateral amygdala lesions (Feinstein, Adolphs, & Tranel, 2016) showed that she has impairment in recognizing fear expressions but was intact in recognizing other expressions including happiness, sadness, and anger. A meta-analysis by Sergerie, Chochol and Armony (2008) described the amygdala’s response in other basic emotions (excluding surprise). They further found that positive emotions (happiness) evoked larger amygdala responses than negative emotions (fear, anger, sadness, and disgust), which is agreeable with Bonnet et al.’s (2015) findings. Comparably, the amygdala’s activity to disgust and contempt was also evident in Sambataro et al.’s research (2006). Wang et al. (2017) recorded 234 single neurons in the amygdala from 9 patients and observed two groups of neurons that were responding to different facial expressions (happiness vs. fear) and emotional categorical ambiguity. Such emotion selectivity neurons in the amygdala was also reported in animal studies (Gothard, Battaglia, Erickson, Spitler, & Amaral, 2007).

The complexity of the amygdala’s role in emotional processing derives from its connection with multiple brain regions (Adolphs, 2002; Armony, 2013). Stimulus information received
via neocortices will be projected back from the amygdala and to other neural structures including the basal forebrain, hippocampus, basal ganglia, hypothalamic and brainstem nuclei to mediate emotional processes and emotional responses (Adolphs, 2002).

1.4.2 Prefrontal Cortices (PFC)

A large body of literature studied the role of PFC in higher-level cognitive processing, including emotional processing (Barbas, 2000; Levesque et al., 2003). PFC is a relatively large region in the brain. With regard to affective processing, three subdivisions are of particular interest, namely dorsolateral, ventromedial, and orbitofrontal sectors (Davidson & Irwin, 1999).

Rapid response to facial expressions and emotions in the PFC was evident in ERP studies. Marinkovic, Trebon, Chauvel and Halgren (2000) recorded the ERP before and after a patient underwent surgical removal of the cortex orbiting the right PFC. They found faces evoked larger ERPs at the PFC area approximately 150ms post-stimulus. After the surgery, the patient depicted severe impairment in facial expression recognition, fear in particular, while the ability to recall emotional words was intact, which can be a strong evidence for the involvement of the PFC in emotional processing. Another study by Kawasaki et al. (2001) looked at single-neuron responses to emotional stimuli in ventral right PFC and reported activation between 120-160ms post-stimulus.

Research on patients with brain damage found that damage to PFC is likely to develop emotional disorders, such as depression and bipolar disorder (BD) (Gainotti, 1972; Lawrence et al., 2004). In particular, Lawrence et al. (2004) observed increased ventral prefrontal cortical activities in BD patients for mild happiness and sadness, and mild and intense fear; while in the control group, they observed increased dorsal prefrontal cortical activities for intense sadness.
Levesque et al. (2003) examined individuals’ ability to self-regulate and suppress negative emotions. They found significant loci of activation in the right dorsolateral prefrontal cortex (DLPFC) and the right orbitofrontal cortex (OFC) when participants were required to suppress emotional response to sad movie clips. Similarly, Goldin et al. (2008) also reported PFC responses in sad emotion suppression conditions.

The synergistic roles of the amygdala and the PFC were found in mediating purposive behaviors (Barbas, 2000). Given the close anatomical connection between the PFC and the amygdala, the PFC may modulate contextually how the amygdala processes facial stimuli as it modulates emotional responses facilitated by the amygdala. The amygdala retrieves affective information from stimuli while the PFC regulates the purposive behavior (Barbas, Ghashghaei, Rempel-Clower, & Xiao, 2002; Levy & Goldman-Rakic, 2000). In addition, the vmPFC is found to play a role in regulating the amygdala activities. For instance, patients with focal and bilateral damage in the vmPFC enhanced the amygdala responses to negative emotions (Motzkin, Philippi, Wolf, Baskaya, & Koenigs, 2015). The increased connectivity between the PFC and the amygdala also resulted in a decline in the amygdala response, which inhibits the increase in cortisol level and thus reduced stress (Gee et al., 2013; Townsend et al., 2013; Veer et al., 2012).

1.4.3 Anterior Cingulate Cortex (ACC) and Insular Cortex

Neuroimaging studies have explored the ACC’s role in emotion and facial expression processing. Killgore and Yurgelun-Todd (2004) used a backward masking of neutral faces immediately after the presentation of happy and sad faces. Masked happy faces activated the anterior cingulate gyrus and amygdala bilaterally and masked sad faces activated only the left anterior cingulate gyrus. Emotion-wise, masked happy faces evoked greater activities in the ACC and amygdala, and masked emotional faces increased the responses in the left ACC and amygdala. They concluded that the ACC and amygdala facilitate the detection and
discrimination of emotions under the conscious threshold. There is also evidence that the
cingulate cortex responded to painful and disgusting video, but disgusting scenes selectively
activated the posterior cingulate cortex (Benuzzi, Lui, Duzzi, Nichelli, & Porro, 2008).

The human insular cortex lies deeply in the lateral sulcus and connects multiple other regions
of the brain (Naidich et al., 2004). In humans, the insula is bidirectionally connected with a
wide range of brain regions, including the frontal, parietal and temporal lobes, the cingulate
gyrus and subcortical regions (Gu, Hof, Friston, & Fan, 2013). Patients with focal brain
damage showing insular lesions were associated with underperformance in facial expression
categorization and recognition (Adolphs et al., 2000). Kipps, Duggins, McCusker and Calder
(2007) studied patients with Huntington’s disease (HD) with structural MRI to map the gray
matter volume change when performing basic emotion recognition task. They found that HD
participants with atrophy in the anteroventral insula were specifically impaired in recognizing
disgust expression relative to other expressions. In a single case, a patient NK with lesions in
the insula, internal capsule, putamen and globus pallidus was selectively impaired in
recognizing disgust expressions. More specifically, this patient would miscategorize disgust
as anger (Calder, Keane, Manes, Antoun, & Young, 2000).

The anatomical connection between the ACC and insula plays an important role in
integrating and executing information, including emotional, cognitive, and autonomic
functions (Taylor, Seminowicz, & Davis, 2009). The research by Taylor et al. (2009)
distinguished two systems in the connection between the ACC and insula. The first system
links the anterior insula (the AIC) with pACC and the MCC, which combines emotional
salience to create subjective emotional experience, particularly negative experience like pain
(Seminowicz & Davis, 2007); while the second system involves the insular cortex and MCC,
which monitors environmental context and response selection. Damasio et al. (2000) also
observed the synergetic interaction between the ACC and insular cortex in processing
emotions, where the anterior pons is deactivated in fear and this region receives projections from the anterior cingulate. When people are in anger and sad status, the regions of the anterior pons and the anterior cingulate activate, and these regions deactivate in happiness. Those regions then receive projections from posterior cingulate and insula (Damasio et al., 2000). In terms of happiness, Damasio et al. (2000) noted increasing activation in the right insula, right secondary somatosensory cortex, left and right anterior cingulate, and right posterior cingulate. Besides, the activations in the right orbitofrontal cortex, left basal forebrain, right hypothalamus, and left midbrain were also observed during happiness.

1.4.4 Other Regions

The superior temporal sulcus (STS) and the fusiform face area (FFA) are the two areas that specialize in early face perception (Adolphs, 2002). The STS is an area of cortex that is believed to control changeable facial features, such as facial expression, eye gaze or lip-speech, while the FFA regulates invariant aspects like identity (Calder & Young, 2005). Yet the findings regarding the proposed role segregation in face perception are inconclusive and less consistent. Few studies found the STS were engaged in facial expression. For example, the study by Narumoto, Okada, Sadato, Fukui, and Yonemura (2001) found that the activity of the right STS increased to facial expressions compared to the face per se. They concluded the special role of the right STS played in facial expression recognition. A MEG study (Stephen et al., 2003) observed activation in the FFA when participants recognized facial expressions of happiness, disgust, and neutral at approximately 150ms. The strongest activation was modulated by happiness, followed by disgust and neutral, respectively.

The activation of anterior temporal lobe neurons is observed in both human and primates with presentation of faces (Perrett et al., 1984), though researchers found that the activity of anterior temporal lobe neurons increased significantly in the monkey responding to facial expressions compared to human (Hasselmo, Rolls, & Baylis, 1989). Ojemann, Ojemann and
Lettich (1992) argued that it might be because neurons that are sensitive to the face found in monkey are embedded somewhere else in the human brain. In Rapcsak, Kaszniak and Rubens’ (1989) study, they reported participants with a right temporal lobe lesion had selective impairment in responding to facial expressions. Additionally, electrical stimulation of the posterior middle temporal gyrus altered emotional labeling of facial expression instead of the perception of faces, which demonstrated the area is specified in facial expression labeling (Fried, Mateer, Ojemann, Wohns, & Fedio, 1982).

A study by Weddell, Miller and Trevarthen (1990) suggested that lesions in the frontal lobe selectively impairs the voluntary production of facial expressions. Coan, Allen and Harmon-Jones (2001) analyzed EEG alpha power when participants produced posed facial expressions. They used the approach/withdrawal model of emotion (Harmon-Jones, 2004). They observed approach emotions (joy and anger) increases the left frontal activities while withdrawal emotions (fear, sadness, and disgust) increases the right frontal activities. The modulation in the frontal lobe by voluntary emotions was also supported by a review study (Phillips, Ladouceur, & Drevets, 2008).

The increasing activity in the right secondary somatosensory cortex for facial expression discrimination tasks is well documented in neuroimaging studies (Vuilleumier & Pourtois, 2007). Right secondary somatosensory cortex was also observed to be activated in happiness and fear, and the entire secondary somatosensory cortex was negatively peaked in sadness and anger (Damasio et al., 2000). Monrad-Krohn (1924) noted several cases of post-encephalitic Parkinsonism which their basal ganglion subcortical structures were affected. He found that their ability to generate spontaneous facial expression was impaired yet the ability to pose expressions was intact. More recent studies (Benuzzi et al., 2018; Borsook, Upadhyay, Chudler, & Becerra, 2010) reported the activation in the basal ganglia when viewing painful facial expressions. In addition to that, voluntary and spontaneous facial
expressions might also be controlled by different neural paths. Clinical studies showed impairment of voluntary facial expressions with cortical lesions and impairment of spontaneous facial expressions with the basal ganglia damage (Rinn, 1984).

1.5 Left and Right Hemisphere Lateralization

Brain asymmetry of emotional stimuli responses is a burgeoning area of emotion researchers. Research indicates functional differences between the two hemispheres in emotional content processing, and the right hemisphere is specialized in processing emotional facial expressions (Narumoto et al., 2001). The right hemisphere superiority in emotional processing is supported by research showing the left side of the face is more active than the right side in expressing emotions (Borod, Haywood, & Koff, 1997) and the right hemisphere is more dominant in facial expression discrimination (Borod, Zgaljardic, Tabert, & Koff, 2001). Neuroimaging and psychophysiological studies found stronger activity in the right hemisphere in perceiving facial expressions (De Winter et al., 2015).

An interesting study combining TMS and EEG found temporal differences of emotional modulation in the right hemisphere, suggesting both hemispheres are involved in emotion discrimination with a temporal advantage in the right hemisphere (Mattavelli, Rosanova, Casali, Papagno, & Romero Lauro, 2016). Further, they found the left hemisphere showed higher activity for neutral faces while the right hemisphere was more active for happy and fearful expressions. In addition, research has also suggested that the right cerebral hemisphere specializes in controlling voluntary facial expressions, while spontaneous facial expressions are controlled by both hemispheres (Ekman et al., 1981).

The inhibition of the right hemisphere tends to lead a decrease in perceiving intensity of emotions displayed on faces (Ahern et al., 1991). When the right hemisphere was
anesthetized in patients, they rated the intensity of facial expressions lower than the baseline ratings. This effect was, however, absent in the anesthetization of the left hemisphere. Likewise, several other studies reported more intensive facial expressions were perceived on the left half of face with non-brain-damaged participants (Borod, Koff, & White, 1983; Dopson, Beckwith, Tucker, & Bullard-Bates, 1984). In addition, the left-sided advantage was more prominent for positive emotions than negative emotions (Borod et al., 1983), and stronger for spontaneous expressions over posed ones (Dopson et al., 1984). On the other hand, some studies revealed inconclusive and non-supportive findings over the right hemisphere dominance model. For instance, Braub, Baribeau, Ethier, Guerette and Proulx (1988) found that negative emotions showed different patterns on face-sided dominance, where sadness was right-face dominant while fear exhibited no hemisphere dominance. Further, they reported that lower face did not exhibit significant left-face dominance, which is in contrast with Thompson’s (1985) study suggesting contralateral control of facial muscle only occurs in the lower face region. Nevertheless, both studies point out that the methodological differences in tasks (elicitation, emotion rating, emotion discrimination, etc.) may contribute to the inconclusive findings of hemosphericity or facial asymmetry (Braub et al., 1988; Thompson, 1985).

Previous studies using a rapid half-filled visual presentation demonstrated that the right and non-dominant hemisphere is responsible for facial processing (Rizzolatti, Umilta, & Berlucchi, 1971; Sackeim, Gur, & Saucy, 1978). In Sackeim, Gur and Saucy’s (1978) research, they observed that facial expressions were more intensive on the left side of the face, which indicated the superiority of right hemisphere in facial expression recognition. Likewise, Philippi, Mehta, Grabowski, Adolphs, and Rudrauf’s research (2009) also corroborated the right hemisphere lateralization effect.
It should be highlighted that studies on the right hemisphere dominance and emotional process did not produce conclusive findings. Emotional process is a large and complex network. Thus some researchers argued that the right hemisphere might be more sensitive to emotional perception and expression (Borod et al., 2001), whereas others were convinced that the right hemisphere is involved specifically in high arousal emotions (Adolphs, Russell, & Tranel, 1999).

1.6 ERP Components of Facial Expressions

The time course of emotional face processing is a historical and ongoing interest in emotion studies. Our brain can rapidly process facial expressions in as short as 40-50ms (Morel, Ponz, Mercier, Vuilleumier, & George, 2009). At approximately 100ms (the P1), a global level of processing emotional versus neutral expression have been observed (Batty & Taylor, 2003). The differences in emotional expressions emerge around 170ms (the N170). The N70 is believed to be evidence of distinguishing faces with non-face objects (Rossion, Joyce, Cottrell, & Tarr, 2003). However, the evidence that whether the N170 is linked to facial expressions are still unsolid. Herrmann et al. (2002) and Eimer, Holmes, and McGlone (2003) both reported that the face-specific N170 was independent and unaffected by facial expressions, which then supported the hypothesis that facial structure perception and facial expression analysis as well as recognition are processed by different pathways in the brain. On the other hand, other researchers (Pourtois, Thut, De Peralta, Michel, & Vuilleumier, 2005) observed larger amplitudes for emotional faces, especially fear expression in comparison to neutral faces. At longer latencies (above 200ms), discrimination between emotional expressions occurs, along with a higher-level cognitive processing, including social emotion perceptions (Batty & Taylor, 2003).
1.7 FACS

Face Action Coding System (FACS) was developed by Ekman and Friesen (1978). It is an anatomically based system to characterize facial behaviors. FACS consists of 44 independent movements, which is termed into Action Units (AUs). Each AU is corresponding to an elemental and independent facial muscle movement and is assigned with a unique AU code. The code represents a single facial movement rather than a single muscle. Although some muscles are involved in different movements, the different movements can be coded differently. For example, the frontalis muscle of the forehead can raise both inner and outer brow, the AU codes for the two movements are AU 1 and AU 2, respectively (Coan & Allen, 2007). On the other hand, some facial movements comprising of more than one muscle will correspond to one AU code only. For example, raising cheeks will contract orbicularis oculi and pars orbitalis, it is identified as a single AU (AU 6) in the system (Coan & Allen, 2007). Taken together, AUs describes functional movements that can possibly appear on the human face (Hwang & Matsumoto, 2016).

FACS can describe a variety of facial behaviors, including but not limited to facial expressions, facial gestures, speech, or chewing. The process of FACS coding requires the identification of facial muscles that involve in visible actions on the face. Furthermore, coding can also encapsulate information of intensity, laterality. and timing of individual AUs. Since all 44 AUs can be independent and comprise additional information, the process is complicated and labor-intensive (Hwang & Matsumoto, 2016).

Several versions are extended from the original FACS to serve different research purposes. In particular, Emotion FACS (EMFACS) were developed by Ekman et al. (1981) to selectively identify AUs that have emotional significance or related to basic emotions rather than coding all 44 AUs (Paul Ekman Group, 2018). BabyFACS was proposed by Oster (2005, p. 261)
because facial expressions of infants are “neither global and diffuse precursors of adult facial expressions, nor precocious, fully formed versions of those expressions”. It works similarly with the adult FACS but has modified some AUs to better superimpose over infants’ facial behaviors. For example, in BabyFACS, AU 3 and AU 4 represent movement of inner corners of the brows and outer corners of the brows, respectively, whereas both movements are categorized under AU 4 in the adult FACS. These differences reflect the structural proportions of infants’ faces, as well as extensive subcutaneous fat deposits and thicker skin (Oster, 2005). Because FACS coding is based on functional anatomy, researchers also extend the coding system to nonhuman primate facial expressions. Studies (Burrows, Waller, Parr, & Bonar, 2006; Liebal, Pika, & Tomasello, 2004; Ueno, Ueno, & Tomonaga, 2004) have shown nonhuman primates and human share structural similarities in facial expression display and even same muscle groups in emotional expressions. Therefore, ChimpFACS was developed to identify the facial movements of chimpanzees (Parr, Waller, Vick, & Bard, 2007; Vick, Waller, Parr, Pasqualini, & Bard, 2007). Apart from chimpanzees, the FACS for orangutans (Caeiro, Waller, Zimmermann, Burrows, & Davila-Ross, 2013) and gibbons (Waller, Lembeck, Kuchenbuch, Burrows, & Liebal, 2012) were also developed to map on facial movements of nonhuman primates.
1.8 Human-Computer Interaction and Human Avatar

The field of human-computer interaction is a proliferating field that draws human cognition and perception to couple with avatars or digital representations for the communication (Ciechanowski, Przeglinska, Magnuski, & Gloor, 2019; Nowak & Fox, 2018). When avatars started sharing extensive similarities, people tended to project their processing strategies of humans to the avatars, which presumes that the avatars are analogous to social others (Nass & Moon, 2000).

To what extent a virtual agent resembles human in terms of biological appearance and behavior influences the connection between individuals and avatars (Sheehan & Sosna, 1991). Neuroimaging studies specified that brain responses differ in perceiving human and non-human agents. For example, watching human cartoon clips activated the mPFC and the cerebellum while these responses were absent when watching non-human cartoon clips (Han, Jiang, Humphreys, Zhou, & Cai, 2005). Even when observing hand actions (grasping actions), only real hands movement activated the right posterior parietal cortex, whereas virtual hand actions elicited responses in lateral and mesial occipital regions.

The term anthropomorphism is often used in avatar research and coupled with human-like or humanoid to describe highly realistic avatars (Ishiguro & Nishio, 2009). It refers to “the attribution of a human form, human characteristics, or human behavior to nonhuman things such as robots, computers and animals” (Bartneck, Kulic, Croft, & Zoghbi, 2009, p. 74). Principally, higher level of anthropomorphism are perceived more attractive and reliable (Westerman, Tamborini, & Bowman, 2015), which in turn contributes to better social communication quality and higher psychological co-presence (Kang & Watt, 2013).

Realism is another term that is commonly used to judge whether an avatar appears vividly human or not. It comprises multiple levels of details, including but not limited to appearance,
facial expression, motion, and even speech (Ishiguro & Nishio, 2009). Similar to anthropomorphism, a higher level of realism aids in a more natural, persuasive, and credible conversation (Park & Chung, 2011).

Highly anthropomorphic and realistic humanoids may fall into the so-called uncanny valley. This was proposed by Mori (1970) who suggested that people increasingly engage with a robot that is more human up to a point at which it is nearly indistinguishable from human, at which point people tend to emotionally reject it. The idea of the uncanny valley will be entailed in Chapter 3.

1.9 The Auckland Face Simulator

The “Auckland Face Simulator” (AFS) is a suite of novel, physiologically based software applications that can be used to create realistic, dynamic, and fully controllable representations of human facial expressions (Sagar, 2015; Mao et al., 2015). It was designed by Mark Sagar and his research team at the Laboratory for Animate Technologies at the University of Auckland as a platform designed to facilitate autonomous human-computer interaction. At the core of the AFS is a computer-controlled dynamic avatar that contains realistic simulations of the facial musculature, postural muscles of the neck, skull, soft tissue, and skin. The AFS offers a tool for stimulus generation that allows the generation of an essentially limitless range of dynamic facial stimuli with an unprecedented level of control.

One of the biggest challenges in creating human-computer interaction agents is to increase the appearance of realism while simultaneously avoiding the “uncanny valley” effect, in which users’ positive feeling of a human-like agent increases along with increasing the human-likeness up to a certain point, at which the image appears very close to human but not yet there. At this point, users will often experience emotional repulsion towards the agent.
(Mori, 1970). I discuss this further in Chapter 3 (see 3.1.1-3.1.5). The trick is not merely pursuing hyper-realistic appearance and movement of the synthetic agents; rather, for a computer-generated agent to look human, it has to react appropriately to different contexts in an emotional and apparently intentional manner (Sagar, 2015). Thus, the AFS includes an integrated system that incorporates both low-level biological details and high-level social interactions. The three core systems integrated in the AFS are (1) a system that simulates the facial musculature underlying expression, (2) a “nervous system” that simulates the cognitive, social and emotional determinants of facial expression, and (3) a social interaction learning system.

1.9.1 Facial Expression and Motor System

A challenge for computer graphics and robotic designers is to create a digital face that avoids emotional disengagement and discomfort to observers. Therefore, Sagar and his team took a variety of factors into consideration. These include physical appearance, coherence of skin movements, and facial expressions (Sagar, Seymour, & Henderson, 2016). The most noticeable feature that distinguishes a real human from an avatar is probably the skin. A lifelike and vivid skin reflects a person’s health status. Visual appearance of the skin can be manipulated or modified by lights, surface, and subsurface details.

Many computer-generated avatars simplify the motion of the skin and model its movement linearly (Sagar et al., 2016). This could result in unnatural facial motions such as a smiling lower face paired with a still upper face. Given the non-linear and interconnected features of facial muscle movements, it is important to treat the entire skin surface of the face system as a single entity, and to model the musculature controlling the deformations of the skin (facial expressions) as an integrated system. Accordingly, the AFS incorporates the FACS (Facial Action Coding System), which was initially developed by Ekman and Friesen (1978) and was subsequently updated in 2002 (Ekman, Friesen, & Hager, 2002). As stated in the earlier
session (see 1.8), producing a facial expression usually involves a combination of facial AUs and miscellaneous actions. AUs are anatomically-based while miscellaneous actions are those without an established anatomic basis. Producing expression using the AFS allows the independent control of each action unit, as well as combinations of several action units to produce specific expressions. The movements of AUs – the contractions of facial muscles – produces realistic deformations of the overlying skin surface, which results in realistic facial expressions, which are linked to established psychophysiology and are specific to the facial structure of each individual avatar. The contractions of individual facial muscles or groups of muscles can be controlled by a user or experimenter, and result in realistic changes to the configuration of the face and produce dynamic and believable changes in facial expression (Sagar, 2015).

In the AFS, simulations of the human face are controlled by a model neural system that incorporates many of the structures known to be involved in human facial expression. The facial nerve (the VII cranial nerve) is predominantly controlled by the motor nucleus in the midpons, which includes motoneurons directed to mimetic muscles, and the facial nucleus in the brainstem (Cattaneo & Pavesi, 2013). The facial nucleus receives projections from five cortical regions that are widely accepted to govern the control of facial expressions, namely the primary motor cortex, the ventral lateral premotor cortex, the supplementary motor area on the medial wall, and the rostral and caudal cingulate cortex (Muri, 2016).

Two different pre-motoneuronal systems govern higher-order and complex facial expressions of emotions. First, facial emotions that respond to threatening stimuli like freezing or escaping stem from the midbrain in the brainstem (Magoun, Atlas, Ingersoll, & Ranson, 1937). Second, the amygdala is known to be involved in fear and fear-related behaviors (Feinstein, Adolphs, Damasio, & Tranel, 2011).
There is evidence that separate neural structures underlie voluntary and emotional facial expressions, and it has been suggested that two parallel systems arise from the facial nucleus (Cattaneo & Pavesi, 2013; Muri, 2016). Muri (2016) has argued that the dissociation between these two systems means that people are not typically able to produce a “genuine” emotional facial expression voluntarily. The AFS includes simulations of both voluntary and emotional facial motor systems (Sagar et al., 2016).

1.9.2 Brain Language

As the facial expression system produces signals with both social and biological significance, to develop a lifelike avatar character, the face should synchronize with both mental and physical states. Inasmuch, building a virtual nervous system that engages real-time learning and sensorimotor interaction becomes the foundation of a hyper-lifelike facial model. Sagar and his team (Sagar et al., 2015) created a software environment called Brain Language (BL) that integrates both “affective and cognitive neuroscience theories” (p. 71). BL is a modular “Lego-like” neural network that includes models of sensory inputs (vision, audition, somatosensation), language processing, emotional response, and so forth. BL links between these modules and the interactive computer graphics to allow realistic interaction with the AFS. This provides a unique platform that allows developing and existing theories from different fields, including but not limited to psychology, neuroscience, and physiology to be integrated into the control of the AFS. This integration allows the output from computations in BL to utilize the realistic computer graphics and animation of the AFS to visualize the outputs from the various BL modules and their interactions.

The main features of BL are autonomy, flexibility, and extensibility. Namely, simple models can be created in BL, and can later be substituted by updated models and/or supplemented by new processing modules (Sagar et al., 2016). This allows users, including researchers, to create and control real-time simulations of the interconnected neural systems that produce
changes in facial expression. The model is comprised of modular computation units or modules, each working like a self-contained black-box, which is expandable at any scale (such as expanding a single neuron to a network). Users can generate and modify algorithmic models in a module. Further, by feeding both spatial and temporal data, the virtual neural network of BL can keep learning. Sensory input (including arbitrary sensory input) to the model can be transferred through camera, microphone, and keyboard, and then translated to the computer graphics output of the AFS. Sagar et al. (2016) underlined that “any variable in the neural network system can be shared and drive any aspect of a sophisticated 3D animation system” (p. 86).

This combination of the realistic animation system that constitutes the AFS with the modular neural control offered by BL is unique. No previous work in facial animation has incorporated virtual models of neural control (Sagar et al., 2015). The models and algorithms that underlie the AFS have been implemented in a number of different platforms. Most prominently, these include “Baby X”, which is an interactive simulation of a human infant, and several models of human adults. In this thesis, I will discuss two such models, “Xyza” and “Leah”, which are simulations of adult women.

1.9.3 BabyX – A Virtual Infant

BabyX is a psychobiological simulation of an infant that integrates interactive behavior and social learning models, which increase the ability of the virtual baby to interact naturally with humans. This model is a cornerstone of the AFS project and is used as a research tool to explore how the neural control of the facial musculature interacts with other aspects of the nervous system. The implemented neural models utilize BL to “create muscle activation based animation as motor output from continuously integrated neural models” (Sagar et al., 2015, p. 80).
**Facial expression and Emotion.** According to Sagar et al. (2016), BabyX’s facial animation is built upon a neuroanatomical architecture that mimics the known facial motor system, whereby signals from motoneuron activities are generated by areas directly or indirectly connected to the facial nucleus. The virtual baby can produce both voluntary and emotional facial expressions, which are created through non-linear modeling of the activation of individual muscles and muscle groups (i.e., action units). To achieve a hyper-realistic character, subtle fine details in the eyes, eyelashes, and teeth are carefully attended.

Biomechanically, the face model is structured with reference to an MRI scan to provide realistic anatomy. Structures like deep and superficial fat, muscles, fascia, and connective tissue are embedded with elasticity to simulate muscle movements from resting to activation.

The emotions of BabyX are established based on a low-level biological system that coordinates and modulates brain-body status. For instance, the simulation of crying is based on a neuroendocrine model of psycho-social stress. Both the amygdala and the hypothalamus control the release of corticotrophin-releasing hormone (CRH), which releases adrenocorticotrophic hormone (ACTH) and cortisol, and triggers BabyX to cry. With BabyX, users can visualize how low-level biological changes (such as changes in the levels of these stress-related hormones) can reflect on and extend to higher-level cognitive behaviors.

**Neural models.** The neural models implemented in BabyX consist of an interconnected virtual neural system and subsystems. With the implementation of Lego-like BL, the models can establish a “large functioning sketch”: an integrated system that interconnects both top-down and bottom-up behavioral mechanisms. To date, the virtual neural systems modeled include “Basal Ganglia, Hippocampus, Hypothalamus, Amygdala, Oculomotor System, Superior Colliculus, Facial Nucleus, and other brain stem nuclei, Cortico-Basal and Cortico-Thalamic Loops, Dopaminergic and other neuromodulatory systems, and higher-level models of episodic and working memory” (Sagar et al., 2015, p. 80). The model also covers basic
neural and behavioral elements, including “model control, behavior selection, reflex, visual attention, learning, salience, emotion and motivation” (Sagar et al., 2015, p. 80) and each neural structure carries particular characteristics.

An example scenario of how the virtual neural models work: when sensory inputs are fed through the camera, the retina detects the luminance change and sends signal to the superior colliculus, which activates the oculomotor system to trigger horizontal eye movements to fixate the region of change in the visual field. Virtual dopamine is released to potentiate motor activities and learning (Sagar et al., 2014). The stimuli trigger the amygdala and hormone release in the hypothalamus. The brainstem motor system will then be activated to generate facial muscle movements. BabyX can also implement neural models to produce subtle behavioral responses, such as pupil dilation and changes in blink rate. Inasmuch, users can modulate virtual neurotransmitter levels to influence BabyX’s behaviors, sensitivities, and even temperament.

**Interactive Learning.** As BL has integrated multiple biologically based learning models, BabyX is able to associate high-level social interactions with low-level biology. We can illustrate this idea by an example of facial mimicry in BabyX. A specialized recurrent neural network that incorporates the basal ganglia thalamic circuit is designed to produce motor babbling, which contributes to a facial animation creation. When a “caregiver” (i.e., a user) mimics the virtual baby’s facial expression, BabyX responds to the sensory input with a phasic release of dopamine that triggers sensorimotor activities to generate active expressions (Sagar et al., 2015). From the perspective of high-level social interaction, it may be easier to simulate a virtual agent by just “transplanting” the facial expressions.
1.9.4 Xyza and Leah

The AFS has expanded beyond the BabyX project to further human-computer interactions with “adult” virtual agents, based on the same computational platform as BabyX (e.g., BL).

In my studies, I employed both Xyza (Figure 1.2) and Leah (Figure 1.3) from the AFS as they were the only two complete models available when I started the experiments. Both models are female characters. Xyza appears to be approximately 22-26 years old, while Leah appears to be in her late 40s. Xyza was initially created as a prototype avatar to fulfill commercial projects by enhancing the engagement of the general public with commercial applications. The model incorporates the autonomous avatar system into the application, allowing a user to interact directly with the avatar. Xyza is designed to have broad appeal to users and attract their attention. She is a “virtual agent” designed to look attractive and sophisticated rather than merely hyper-realistic.

In contrast to Xyza, whose main purpose is commercial and consequently to draw users’ attention and interact with them, Leah was developed as part of a research project investigating psychobiological virtual agents. As a result, she was created with imperfections such as age spots and wrinkles that remarkably make her appear more human.
Figure 1.2 Xyza

Source: The University of Auckland (2019)
Figure 1.3 Leah

Source: Lawler-Dormer (2017)
In principle, the AFS aims to give precise control to facial expression and emotional studies. Therefore, it is important that we can get to the point where the AFS-produced stimuli are as good as human or photo stimuli. This is very similar to the “Turing Test” in the artificial intelligence world where machines perform equally to human. Despite advances in computer science and technology, engineers and scientists are not able to create a machine that can think like a human or pass the Turing Test. The idea of Turing Test, can however be transplanted to the development of realistic face avatars. The ultimate goal is to create a computer avatar that can pass the facial version of the Turing Test.

Turing Test, was inspired by Alan Turing, who proposed to consider the question – “Can machines think?” in 1950 (Turing, 2009). To detangle this problem, Turing has exemplified using a game called the “imitation game”. In the game, there will be three players, a man, a woman, and an interrogator. The goal of the game is for the interrogator to identify which of the other two players is the man and the woman respectively. The interrogator is provided labels of X and Y to represent the other two players. Ultimately, he or she will determine “X is the man, Y is the woman” or “X is the women and Y is the man”. So, if a machine takes part in the man’s role, will the interrogator play the same as he or she does when the game involves only human? When the interrogator’s performance does not distinguish between a human player and a machine player, we will say the machine has passed the Turing Test. Based on that, a few years later, the concept of artificial intelligence was then developed and coined by John McCarthy in 1956, who is now known as the father of AI (Sharkey, 2012).

To date, there are movie or animated characters (e.g., Beowulf in the movie Beowulf 2007) that may be visually indistinguishable from human. However, our brain may still detect tiny differences between an avatar and a photo, which could signify the manifestation of the uncanny valley.
1.10 Research Questions and Hypothesis

The main question that motivated this thesis is whether faces simulated using the AFS can be perceived like real human faces. In other words, could the AFS pass the facial Turing Test - or even come close to passing it?

To address this question, I started with the influence of facial expression of emotion. Firstly, I investigated the threshold intensity at which an emotional expression could be perceived by an observer. Therefore, I selected lower intensity expressions as stimuli. In everyday life, people do not normally wear full expressions. Instead, partial or subtle expressions are more frequent. I used ERPs to explore whether observers are sensitive to different emotions expressed at lower intensity, specifically 20% and 40% of the “full” expression. This question was explored in Chapter 2.

In Chapter 3, I looked at the ERP responses to photo and the AFS faces. In particular, I examined whether the brain could distinguish between photo and simulation and compared behavioral ratings between the two categories of stimuli. Additionally, I also explored the impact of emotions and intensity levels on people’s perception of virtual agents’ realism.

At the time of the experiment, I had two AFS models, Xyza and Leah, both of which are simulations of adult females. Xyza looks to be in her early 20s, and Leah appears to be in her late 40s. I was then intrigued to know whether I can use ERPs to identify responses to subtle differences in realism between these avatars, which may open an opportunity for psychophysiological testing of human-likeness between avatars other than relying on simple eyeballing. The analysis was included in Chapter 4.

In Chapter 5, I aimed to explore the underlying possibilities that could explain ERP differences between photo and simulated faces in Chapter 3 from the perceivers’ side. I
assumed that the differences may derive from how perceivers compare photo and simulated faces in general.

In Chapter 6, I have presented the main findings and contributions of this research. I further discussed the limitations of my studies and explored possible future work in using hyper-realistic avatars in research.
Chapter 2 Identifying Subtle Expressions

2.1 Introduction

Do you still remember when you were punished in school, and you had to talk to your mother about it? You would cautiously look for cues on her face, from tiny twist of the mouth to more obvious frowning. Have you ever been betrayed by your facial expressions - perhaps at a social event when you see someone with funny hair? You are supposed to suppress your emotion, yet end up laughing or smirking. A subtle change can sometimes create a substantial difference in the meaning we derive from facial expressions. For instance, we may be uncertain if a person is actually happy when seeing a quick and brief smile. On the other hand, in daily life, we can retrieve so much information from a facial expression almost effortlessly. It is often said that a simple expression is worth more than a thousand words. The dynamics and complexity of these expressions are what attracts researchers into the world of facial expressions of emotion.

2.1.1 Facial Expression of Emotion Processing

Emotions are an inevitable and important aspect of human communication. Facial expressions act as a doorway to access emotional cues without the need for verbalizing; a smile is sufficient to convey happiness without the need for words.

Bruce and Young’s (1986) influential parallel-processing model of face and facial-expression processing suggests that the different aspects of a face, including identity and facial expressions, are processed independently. They proposed that a face contains seven types of information, namely pictorial, structural, visually derived semantic, identity-specific semantic, name, expression, and facial speech codes.

According to Bruce and Young, the pictorial code of a face incorporates details like lighting, grain, imperfection, the portrayed pose and expressions of a picture. In the laboratory, the
pictorial code can facilitate episodic memory for faces in yes/no recognition tasks because unfamiliar faces are presented and tested using the same pictures. In everyday life, however, pictorial codes are less important, since faces are seldom encountered with identical details – in other words, identical pictorial codes. The structural code captures the configurational information required to discriminate a face from other faces. Bruce and Young further specified that the pictorial code responds to any visual pattern or picture, while the structural code mediates the differentiation of a face from low spatial frequencies or caricatures when pictorial codes are not identical. The visually derived semantic code contains information about a face that can be shared between familiar and unfamiliar faces, including age, sex, ethnic origin, honesty or intelligence. The identity-specific semantic code goes beyond physical features to specify semantic aspects of an individual face. The relationship between visually derived and identity-specific semantic codes can be analogous to “the semantics of a word in relation to its spelling” (Bruce & Young, 1986, p. 309). The name code simply specifies a person’s name. Bruce and Young argued that it could be regarded as a particular type of the identity-specific semantic code since it is possible for a person to have an identity-specific semantic code without a name code. Expression and facial speech codes generally take the literal meaning wherein expression code specifies facial expressions of emotion, and facial speech code describes movements of the lips and tongue when talking. They further argued that these two codes do not serve any importance in the face recognition.

Global structural codes, combined with view-centered information, lay a foundation for the analysis of facial expression and speech. Bruce and Young’s (1986) model includes “face recognition units (FRUs)” that process identity independently of expression. When recognizing a face, the expression-independent structural description is transferred to the FRUs to process further (Posamentier & Abdi, 2003). The model therefore suggests that identity and facial expression processing occur in parallel (see Figure 2.1).
Studies of patients suffering prosopagnosia offer the most striking support for the dissociation between facial identity and expression processing. Depending on the severity, prosopagnosic patients can fail to recognize the faces of their close family or even themselves (Wacholtz, 1996). Despite the severity of this “face-blindness”, the recognition and reading of basic facial expressions is preserved in at least some patients (Lee, Duchaine, Wilson, & Nakayama, 2010). Nevertheless, brain damage can affect a large area, and it is arguable that prosopagnosia only influences face processing. Some prosopagnosic participants also show deficits in recognizing objects (Gauthier, Behrmann, & Tarr, 1999). Thus Gauthier et al.
(1999) argued that the deficit in prosopagnosic patients is better defined as a deficit in within-category discrimination rather than face-specific processing. As a result, the evidence from prosopagnosia patients in advocating the dual-routes face processing may not be as concrete as it sounds. However, behavioral studies with normal participants found reaction times were shorter for familiar faces in identity matching task whereas no difference in reaction times between familiar and unfamiliar faces was found in facial expression matching task (Young, McWeeny, Hay, & Ellis, 1986), again supporting the parallel routes of processing between emotion and identity.

2.1.2 Subtle vs Overt Facial Expressions

People express emotions or feelings in various ways, which can generally be grouped into two major categories: overt and covert channels. Language or verbal communication is a straightforward and explicit way to deliver emotions, such as saying “I am frustrated”, “I am happy”, or “this smells disgusting”. Moreover, language provides specific words to describe one’s feelings, like happiness, sadness, anger, and others (Enfield & Wierzbicka, 2002).

Besides verbal language, people can express emotions through gestures, posture, and other aspects of “body language”. Like verbal language, body language can convey an emotional state explicitly too; clenching a fist is an explicit indication of anger, for example.

Nonhuman animals also display their emotion by body language. A dog will run in circles and towards the door, if he/she senses that you are going to take him/her for a walk (Sanders & Arluke, 1993).

However, the overt channels are not the only means of communicating emotional information. Sometimes people try to conceal their emotions by controlling body languages and talking in carefully selected words and sentences. In such cases, emotional states are frequently betrayed by more covert means, such as by facial expressions.
According to the Paul Ekman Group, there are three types of facial expressions, namely macro-, micro- and subtle expressions (Paul Ekman Group, 2016). What differentiates between a micro and a macro expression is fundamentally the time duration. Macroexpressions usually last between half second and 4 seconds, whereas it agreed on a general level that expressions last less than half a second would be categorized as microexpressions (Paul Ekman Group, 2016). Subtle expressions are facial expressions with lower intensity and involved fewer Action Units [as defined in the Facial Action Coding System (FACS; Ekman & Friesen, 1976)] (Matsumoto & Hwang, 2014). Unlike microexpressions, subtle expressions have no time limits and can be either a low intensity full expression or a fragment of a full expression (Matsumoto & Hwang, 2014).

It appears that people are not particularly good at observing or recognizing subtle or microexpressions (Frank, Herbasz, Sinuk, Keller, & Nolan, 2009), even though some studies suggest that recognition accuracy can be improved with training (Matsumoto & Hwang, 2011). Both subtle and microexpressions involve low facial expression intensity, which then raises the question: to what extent we are capable of differentiating different intensities of a facial expression?

For an artist, the intensity of a particular expression can convey different meanings. Questions such as whether lowering the eyebrow makes one look angrier, or if widening the mouth open appears to be more surprised are often discussed in artwork production or evaluation (Faigin, 1990). The question can be further developed into how people perceive facial expressions with different intensities, and whether the intensity of expressions affects the perception of an emotion.

The discrete-emotions model argues that distinct expressions activate different intensity levels of facial muscles to portray specific emotions (Ekman, 1993), whereas dimensional
emotion models argue that valence and arousal combine to index intensity and discrimination of different facial expressions (Carroll & Russell, 1996).

Surguladze and colleagues (2004) examined the ability of patients with Major Depressive Disorder (MDD) to identify subtle expressions. Happy and sad expressions that were morphed at 50% intensity and full intensity (100%) were used as the stimuli. They found that in comparison to full intensity happiness, individuals with MDD were less likely to correctly identify 50% happy expressions. Along the same lines, Joomann and Gotlib (2006) also reported different intensity levels of expressions required to recognize emotions in individuals with MDD and social phobia. Participants with MDD required a higher intensity to identify happiness and lower intensity for anger, whilst individuals with social phobia require less intensity to recognize anger.

Hess, Blairy and Kleck (1997) also raised the question of decoding accuracy for expressions with different intensity levels. They used four expressions as stimuli: anger, disgust, happiness and sadness, and concluded that “decoding accuracy varied largely with the physical intensity of the expressions for all emotions except happiness” (Hess et al., 1997, p. 253). That is, their results revealed a general increase in decoding accuracy along with the increase in emotional intensity level. An exception is happiness, which remained close to 100% accuracy at all intensity levels. The findings are consistent with Palermo and Coltheart (2004), who reported a high accuracy rate of happy expression recognition at across different intensities.

An understanding of the influence of facial expression intensity is also important for the creation of robotic or humanoid agents. Makarainen, Katsyri and Takala (2014) investigated the relationship between facial expression intensities with different levels of realism (from schematic to photo faces) and perceived strangeness of the faces. They found that the
perceived intensity of facial expressions decreased when the degree of realism decreased. To compensate, the intensity of the expressions may be exaggerated. In other words, when there is a virtual face with a lower level of realism, for its facial expression to reach to the same level of intensity as perceived on a real human face, the expression tends to be exaggerated compared to the real human expression.

2.1.3 ERP and facial expression

Event-related brain potentials (ERPs) offer the possibility of determining the time course of emotional processing from facial expression. ERP studies have shown that neural activation related to unconscious facial expression perception can arise within the first 100 to 200ms, (Molholm, et al., 2002). Some studies (Di Russo, Martinez, Sereno, Pitzalis, & Hillyard, 2002; Hinojosa, Mercado, & Carretie, 2015; Itier & Taylor, 2002) suggested that the early processing of facial expressions is usually pre-attentive and automatic at a more holistic and structural level. These studies have investigated emotional expression modulation of early ERP components within the initial 200ms post-stimulus, including the P1 and the N170. By contrast, conscious perception of facial expressions usually occurs beyond 200ms post-stimulus, which signifies a finer process of visual sensory cues and top-down feedback (LeDoux, 1998; Kiefer & Spitzer, 2000). For example, Williams and colleagues (2004) employed ERP to compare non-conscious discrimination and detection (stimuli presented for 30ms and 10ms respectively) and supra-threshold detection (stimuli presented for 170ms) of fearful face images. They found that unconscious perception of fear elicited a greater amplitude of the N2 component at about 200ms after stimuli presentation, whereas conscious perception of fear elicited a larger negativity at around 400ms. The N170 and N2 is reported to index the holistic analysis of facial configurations. The modulation may be attributable to the distinction of configurational details between facial expressions (Bentin, Allison, Puce, Perez, & McCarthy, 1996). This is also in light with previous ERP studies (Bernat, Bunce, &
where affective responses could occur with the absence of consciousness, and the process starts emerging within 200ms. Conscious process of emotional contents for both face and non-face stimuli commenced around 400ms (Kiefer & Spitzer, 2000).

In the current study, I focused primarily on early ERP components, including the P1 and N170. I also included an emotion-related modulation of the ERP recorded at fronto-central electrodes, termed the “Frontocentral Emotional Positivity” (FcEP) which has a broad temporal distribution from approximately 200ms to 400ms (Hilimire, Mienaltowski, Blanchard-Fields, & Corballis, 2014).

2.1.3.1 P1

Although the N170 is widely accepted as an early face-specific ERP component, whether it is the earliest component influenced by face processing remains in dispute. The P1, which is usually maximal at parietal and/or occipital electrodes and peaks around 80 to 130ms, is thought to primarily reflect low level visual processing (Batty & Taylor, 2003; Di Russo et al., 2002). This component may also arguably be modulated by higher-order cognitive processing of faces. For instance, Pizzagalli, Regard and Lehmann (1999) demonstrated brain potentials as early as 100ms after stimulus onset responded to liked and disliked faces. They concluded that the P1 component is probably the earliest component that shows valence-dependent modulation by emotions, which could, in turn, be a manifestation of pre-attentive and automatic emotional processing. Similarly, Meeren, van Heijnsbergen and de Gelder (2005) employed face-body compound images with fearful and angry expressions that were either congruent or incongruent with body gestures. They found that congruent conditions elicited an enhanced amplitude of the P1 for incongruent conditions, which they attributed to an automatic and holistic extraction of affective information. Other studies (Itier & Taylor, 2004; Itier & Taylor, 2002) also found a shorter P1 latency for upright faces, and a longer
latency and larger amplitude for inverted faces than upright faces. These effects are consistent with the suggestion that face processing is prioritized, so that global information about faces is extracted in the initial 100ms (Sugase, Yamane, Ueno, & Kawano, 1999; Taylor, 2002).

2.1.3.2 N170

The N170 is known to be elicited by face stimuli compared to other objects and is typically interpreted as a face-specific component (Bentin et al., 2007; Eimer, 2000). It is thought to index early visual processing for the configuration of a human face. Whether different emotional expressions also modulate the N170 component remains in debate.

Early studies on the N170 majorly showed no effect of facial expressions. Eimer and Holmes’ (2002) study looked into the time course of facial expression processing. They found that the frontocentral emotional positivity (FcEP) could be activated as early as 120ms for an upright fearful expression which elicits a higher-amplitude FcEP than neutral faces. However, they also reported that the N170 was not modulated by facial expressions (Eimer, Holmes, & McGlone, 2003). Similarly, Herrmann and colleagues (2002) compared the N170 between happy, sad, and neutral expressions, and observed no difference in either amplitude or latency. Thus, they argued that at early stage, brain activities of different facial expressions are not associated with face perception. Facial expression processing may arise at a later latency (Krolak-Salmon, Fischer, Vighetto, & Mauguiere, 2001).

However, several lines of evidence have demonstrated that emotional facial expressions can influence early components like the N170. Batty and Taylor (2003) used both facial and nonfacial stimuli and instructed participants to respond to non-facial stimuli like cars, butterflies, or planes. Their findings showed modulation by facial expressions on both amplitudes and latencies for both the P1 and the N170 components. The results were in agreement with Sugase et al. (1999), who presented monkey and human faces with different facial
expressions to macaque monkeys and recorded single-neuron activities in the temporal cortex. They further found that a finer process reflecting the analysis of expression or identity occurred after the initial global facial categorization, which is supported by the neuron transmission rate of global and finer facial information peaking at approximately 117 and 165ms respectively, which corresponds to the time window of the P1 and N170 components.

A meta-analysis by Hinojosa et al. (2015) reported the modulation of amplitude for the N170 was observed when comparing emotional expressions with neutral expression. Moreover, their findings revealed that the component tends to be sensitive to certain expressions, specifically anger, fear and happiness (Bublatzky, Gerdes, White, Riemer, & Alpers, 2014; Tortosa, Lupianez, & Ruz, 2013). Therefore, Hinojosa and colleagues concluded that the N170 may involve both structural encoding and facial features (including facial expressions and identity) processing in parallel. They further indicated that stimulus variation (such as luminance, size, or contrast) may contribute to the inconsistent results of facial expression modulation in the N170.

Another line of evidence of emotional modulation of the N170 comes from schizophrenia patients, who often have impaired ability in identifying faces and recognizing facial expressions (Barkhof, de Sonneville, Meijer, & de Haan, 2015; Savla, Vella, Armstrong, Penn, & Twamley, 2012). Shah et al. (2018) tested schizophrenia patients with emotional recognition task using four facial expression stimuli (joyful, angry, fearful and sad), a neutral emotional stimulus and a non-face stimuli (a chair) as the control condition. They observed that schizophrenia patients produced smaller N170 amplitudes for all emotional stimuli (including neutral) than the control group. In addition, N170 amplitude modulation by neutral, joyful, angry, and fearful expressions was significant in the control group but not in patients with schizophrenia. Collectively, the difference of facial expression modulations in
N170 amplitudes between the two groups further substantiate the response of facial expressions may start as early as the N170 time window.

2.1.3.3 \textit{Fronto-central Emotional Positivity (FcEP)}

Several studies found emotional expressions elicit more positive ERPs than neutral faces in frontocentral areas (Hilimire et al., 2013). Eimer and Holmes (2002; 2007) reported a more positive amplitude was triggered at an early phase (around 120ms after the onset of stimuli) by emotional faces compared to neutral faces. They termed the ERP component the fronto-central emotional positivity (FcEP), which suggests that emotional expressions can be recognized and processed at a fairly early stage by the prefrontal cortex. The modulation of the FcEP amplitude is shown in three time windows (120ms, 180ms and 250ms post-stimulus), but may contain two separate stages (Eimer & Holmes, 2007). The early emotional expression effect (onset within 200ms post-stimulus) was even present with face recognition tasks where emotions were extraneous, and such effects exhibited a frontocentral scalp distribution. The inverted fearful faces also elicited early frontocentral activities although the component was delayed (onset about 150ms post-stimulus) and attenuated. Besides fear, other facial expressions (including anger, happiness, disgust, surprise, and sadness) showed enhanced positivity against neutral expressions at electrode Fz starting approximately 180ms and sustaining throughout the entire epoch duration (1000ms), although the amplitude difference within the emotional expressions did not reach the statistical significance. The similarity of the ERP effect for the six expressions indicates that emotional expressions were processed in a similar manner, at least within the first 200ms (Eimer et al., 2003). The later stage (250ms post-stimulus and beyond) has a more sustained and broad positive distribution over the scalp for fearful faces, both upright and inverted (delayed and attenuated response). The activities indexed higher-order and more thorough conscious processing of emotional
cues, not limited to facial expressions, but emotional non-face stimuli as well (Keil et al., 2002).

2.1.4 The current study

A majority of studies have investigated the recognition of overt or macro facial expressions with respect to different emotions using psychophysiological techniques. However, facial expressions are multi-dimensional stimuli that differ in emotional contents and intensities of the emotion. They also differ in how prominently an expression is displayed (Sprengelmeyer & Jentzsch, 2006). To date, very few studies have explored ERP responses to more-subtle, lower-intensity expressions. Consequently, the mechanisms of perceiving different intensities of expressions and to what extent individuals can realize the difference of expression intensities remained unclear. The existing ERP literature has barely explored the topic of facial expression intensity, particularly for lower intensity or subtle expressions. To address this issue, I employed ERPs to map the perception of lower levels (intensities) of facial expressions (20% and 40% of the full facial expression). The temporal resolution of ERPs enables studies of the time-course of expressions as well as of the amplitude modulation by expressional stimuli. Here, I explore the psychophysiological “thresholds” for the response to low-intensity emotion.

In the present ERP study, participants were asked to view different expressions (anger, happiness, neutral and disgust) with different intensities (for anger and happiness, the intensity levels are 20% and 40%; for disgust, the intensity level is 100%; and no intensity level for neutral – 0%). Participants were required to do a task-irrelevant response to disgust expressions, which the trials were not analyzed. Therefore, participants were semi-attended to the target expressions – happiness, anger, and neutral.
I hypothesized that ERP components would not differ much between neutral expressions and 20% intensity across the facial expressions, as I presume that individuals can barely see 20% intensity expressions as “an expression”. I further hypothesize that the recognition of “an expression is there” occurs between 20% and 40% intensity, and will be signaled by modulations of ERP components.
2.2 Methodology

2.2.1 Participants

18 participants (11 females and 7 males) were recruited via a mailing list and the SONA system (in which participants register to participate in research studies for course credits in the School of Psychology, The University of Auckland), aged from 19 to 46 years ($M = 24.76, SD = 7.67$). Participants had normal or corrected-to-normal vision, and reported no history of epilepsy or migraine. Moreover, those who had difficulties in recognizing faces or emotional expressions were excluded from analysis. They were compensated with either $20 vouchers or course credits. The study was approved by the University of Auckland Human Participants Ethics Committee.

2.2.2 Materials

The study used the Montreal Set of Facial Displays of Emotion (Beaupre & Hess, 2005), a set of morphed faces with a whole range of different expressions at different intensities from 0% to 100%. Three expressions were selected from two Caucasian female and two Caucasian male models: happy, angry (both at 20% and 40% morphing levels) and neutral. The study employed an oddball paradigm, consisting of 600 trials, which were divided into 4 blocks of 150 trials each. Within each block, “standard trials” (60%) presented a neutral face (coded as 0% intensity of expression), “target trials” (10%) in which a disgusted face (coded as 100% intensity of expression) was presented, and “oddball trials” (30%) which were equally divided between happy and angry faces. Happy and angry faces were further split into 20% and 40% intensity morphs of expressions. The order of appearance of trials was randomized in each block, which lasted for about 10 minutes.
2.2.3 Procedure

Testing occurred in a dimly lit, sound attenuated, and electrically shielded cabin. Participants sat in front of a computer screen with a viewing distance of 57cm. The stimuli appeared centrally on a white background. Each participant completed all 600 trials (approximately 45 mins duration), and all trials were randomized. Each trial started with a black central fixation cross on a white background for 500ms, followed by the face stimulus for 300ms. An inter-trial interval of 3 seconds followed each trial. Participants were required to press the space bar when they saw a disgust facial expression on the screen, and do nothing if they saw other expressions (e.g., anger). Figure 2.2 illustrates the experimental procedure.
Figure 2.2 Trial procedure and the stimuli of the experiment
2.2.4 EEG Recordings

The electroencephalogram (EEG) was recorded from 128 electrodes using an Electrical Geodesics (EGI) Net Amps 400 amplifier. All impedances were kept below 4KΩ. EEG activity was collected with reference to Cz and was re-referenced off-line to an average reference. All channels were resampled at 250Hz and filtered with a passband of 0.10-30Hz. The continuous EEG data was then segmented into 800ms epochs, starting 200ms before stimulus onset, which was also served as a baseline. Blinks and horizontal ocular activities were removed by applying independent component analyses (ICA). Epochs containing amplitudes exceeding +100 or -100μV were also removed. Linear detrending was then applied to the remaining epochs.
2.3 Result

Grand averages and scalp distributions across all conditions are presented in Figure 2.3. Inspection of these waveforms revealed several components of interest. The first of these is the P1 (85-110ms), with a bilateral posterior scalp distribution, which is maximal at the P7 (electrodes including E58, E59, E65, E64, E63, E57, E50 and E51) and P8 clusters (electrodes including E96, E100, E99, E95, E90, E91, E97 and E101). This is followed by a dipolar complex, with negativity broadly distributed across the posterior scalp (N170: 120-160 ms), which I took the measurement at the P7 and P8 clusters (Su, Chen, He, & Fang, 2012), and positivity focused around the Fz cluster [this corresponds to the early FcEP reported by Eimer and Holmes (2002), and Hilimire et al. (2014)] (electrodes including E11, E10, E4, E5, E12, E19, E18 and E16). A wide, dipolar complex follows, with positive voltages at posterior electrodes, and negative at frontal sites from 210-380ms (the timing of this complex coincides with the late FcEP, reported by Eimer and Holmes (2002), and Hilimire et al. (2013), but is more broadly distributed). I also selected the time windows and electrode clusters based on these grand average waveforms and scalp distributions. Because the late complex is broadly distributed in time, I analyzed the first half as an earlier phase (210-295ms) independently of the second half as a later phase (295-380ms). I termed the two phases FcEP Late I and FcEP Late II. It is also worth noting that the recording epoch includes responses to the text in the blinking slide that appeared after the offset of the stimuli (300ms), so I did not analyze beyond 385ms.

For each of the components identified above, I measured amplitude by taking the average voltage in the time windows specified. Table 2.1 displays the time window and electrode clusters specified for each component. Figure 2.3 shows the grand average waves collapsed across conditions.
Table 2.1

Details of Components Analyzed

<table>
<thead>
<tr>
<th>Component</th>
<th>Time window</th>
<th>Electrode clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>85-110ms</td>
<td>P7 and P8 clusters at posterior sites</td>
</tr>
<tr>
<td>N170</td>
<td>120-160ms</td>
<td>P7 and P8 clusters at posterior sites</td>
</tr>
<tr>
<td>FcEP Early</td>
<td>120-160ms</td>
<td>Fz cluster at frontal site</td>
</tr>
<tr>
<td>FcEP Late I</td>
<td>210-295ms</td>
<td>Fz cluster at frontal site</td>
</tr>
<tr>
<td>FcEP Late II</td>
<td>295-380ms</td>
<td>Fz cluster at frontal site</td>
</tr>
</tbody>
</table>

I analyzed both N170 and FcEP early components depicting the dipolar structure in the same time window because I see the possibility that different responses were generating at both frontal and posterior sites.
Figure 2.3 Grand average waves collapsed across emotions (happiness and anger) and intensities (0%, 20% and 40%). Shaded areas represent ERP components of interest (in temporal order - Fz: the FcEP Early, the FcEP Late I&II; P7 & P8: the P1 and the N170)
For the visual posterior components, the P1 and the N170, I ran a 2 x 2 x 3 repeated-measures ANOVA with the amplitude as the dependent variable to explore the influence of emotion and intensity on these components, and location (left, P7 and right, P8 cluster), emotion (anger and happiness) and intensity (0%, 20%, and 40%) as factors. To specify, 0% expressions are neutral expressions. In order to fill out the design of the ANOVA, I randomly assigned each trial for the neutral stimuli to one of the two emotions.

**P1 component.** Participants generated a significantly larger P1 at the right hemisphere $F(1, 16) = 6.320, p = 0.023, \eta^2_p = 0.283$ compared to the left hemisphere (0.497 µV vs. 0.239 µV). Figure 2.4 and Figure 2.5 displayed topographic maps and waveforms for anger and happiness respectively. Neither emotion ($F(1, 16) = 2.718, p = 0.119, \eta^2_p = 0.145$) nor intensity ($F(2, 32) = 1.172, p = 0.317, \eta^2_p = 0.068$) showed any significant effect (see Figure 2.6).
Figure 2.4 a) Scalp distributions for the P1 component (85 – 110ms) of the anger condition. Topographic maps from left to right in order are 0%, 20% and 40% angry expression respectively. b) Waveforms from left to right corresponding to P7 and P8 electrode clusters chosen from posterior sites. The dashed rectangles indicate the magnified P1 component from the waveform for better visualization.
Figure 2.5 a) Scalp distributions for the P1 component (85 – 110ms) of the happiness condition. Topographic maps from left to right in order are 0%, 20% and 40% happy expression respectively. b) Waveforms from left to right corresponding to P7 and P8 electrode clusters chosen from posterior sites. The dashed rectangles indicate the magnified P1 component from the waveform for better visualization.
Figure 2.6 Bars show the amplitude of the P1 component for each expression condition at left (P7 cluster) and right (P8 cluster) scalp locations. The curved bracket illustrates the significant main effect between left and right electrode clusters. *p < 0.05. Error bars show SEM.

**N170 component.** The topographic maps and waveforms of anger and happiness for the three intensity levels of the N170 is displayed in Figure 2.7 and Figure 2.8. There was a significant main effect of emotion on the N170 amplitude, $F(1, 16) = 12.108, p = 0.003, \eta^2_p = 0.431$. It suggested that anger triggered a larger N170 compared to happiness (-0.462 μV vs. -0.374 μV). An interaction between emotion and intensity also reached significance, $F(1, 16) = 6.752, p = 0.004, \eta^2_p = 0.297$. Follow-up pairwise comparisons with a Bonferroni adjustment were performed and showed that N170 amplitude differed significantly when two emotions achieved 20% [anger (-0.439 μV) vs happiness (-0.354 μV), $p = 0.001$] and 40% intensities [anger (-0.501 μV) vs happiness (-0.336 μV), $p = 0.027$]. Anger and happiness showed no systematical differences at 0%, which is expected as they were actually neutral expressions. Furthermore, no enough evidence suggested there were differences across three intensity levels for anger, whereas 40% happiness differed significantly from 0% (-0.336 μV vs. -0.431 μV), $p = 0.043$. Amplitude differences between each condition are shown in Figure 2.9.
Figure 2.7 a) Scalp distributions for the N170 component (120 – 160ms) of the anger condition. Topographic maps from left to right in order are 0%, 20% and 40% angry expression respectively. b) Waveforms from left to right corresponding to P7 and P8 electrode clusters chosen from posterior sites. The dashed rectangles indicate the magnified N170 component from the waveform for better visualization.
Figure 2.8 a) Scalp distributions for the N170 component (120 – 160ms) of the happy condition. Topographic maps from left to right in order are 0%, 20% and 40% happy expression respectively. b) Waveforms from left to right corresponding to P7 and P8 electrode clusters chosen from posterior sites. The dashed rectangles indicate the magnified N170 component from the waveform for better visualization.
**FcEP Early component.** For the mean amplitude of early FcEP early, I found a significant main effect of emotion, $F(1, 16) = 8.224, p = 0.011, \eta^2_p = 0.339$. Again, anger elicited a more positive FcEP amplitude than happiness (0.525µV vs. 0.388µV). A significant interaction of emotion and intensity was revealed, $F(1, 16) = 3.639, p = 0.041, \eta^2_p = 0.185$. Pairwise comparison with Bonferroni correction showed FcEP amplitude was larger for anger than happiness at 20% intensity [anger (0.590 µV) vs happiness (0.336 µV), $p = 0.011$]. At the 40% level, anger also evoked a larger amplitude, but only marginally reached the significance threshold ($p = 0.059$). I failed to observe significant amplitude differences across intensities within each emotion, all $ps > 0.05$. This interaction, along with the relevant waveforms and topographic maps, is displayed in Figure 2.10.
Figure 2.10  a) Scalp distributions for the FcEP early component (120 – 160ms) of anger and happiness condition. Topographic maps from left to right in order are 0%, 20% and 40% for angry and happy expression respectively.  
b) Waveforms from left to right corresponding to Fz electrode clusters chosen from anterior sites for anger and happy condition. The dashed rectangles indicate the magnified FcEP early component from the waveform for better visualization.  
c) Bars show the amplitude of the FcEP early component for each expression. The curved bracket illustrates the significant main effect between anger and happiness emotions. *p < 0.05. 
Error bars show SEM.
**FcEP Late I component.** It revealed a marginally significant interaction effect between emotion and intensity, $F(1, 16) = 3.409, p = 0.059, \eta^2_p = 0.176$. Follow-up pairwise comparison, with Bonferroni correction, suggested that anger generally triggered more positive amplitude than happiness at the 20% level, but the differences failed to reach a statistically significant level, all $p$s > 0.05. Figure 2.11 depicts the scalp distribution, waveforms, and amplitude across the conditions.
Figure 2.11 a) Scalp distributions for the FcEP late I component (210 – 295ms) of anger and happiness condition. Topographic maps from left to right in order are 0%, 20% and 40% for angry and happy expression respectively. b) Waveforms from left to right corresponding to Fz electrode clusters chosen from anterior sites for angry and happy condition. The dashed rectangles indicate the magnified FcEP late I component from the waveform for better visualization. c) Bars show the amplitude of the FcEP late I component for each expression. Error bars show SEM.
FcEP Late II component. No enough evidence for significant main effects or interactions were found for this component, all $p$s > 0.05. The topographic maps, waveforms, and bar graph are shown in Figure 2.12.

Figure 2.12 a) Scalp distributions for the FcEP late II component (295 – 380ms) of anger and happiness condition. Topographic maps from left to right in order are 0%, 20% and 40% for angry and happy expression respectively. b) Waveforms from left to right corresponding to Fz electrode clusters chosen from anterior sites for angry and happy condition. c) Bars show the amplitude of the FcEP late II component for each expression. Error bars show SEM.
2.4 Discussion

In this study, I aimed to investigate the threshold for evoking ERP responses to emotional facial expressions. I measured the ERPs evoked by happy and angry facial expressions at 20% and 40% level, as well as neutral expression (0%). Questions of interests are: a) the just noticeable difference (JND) intensities of facial expressions of emotion against each other; and b) the required intensity of recognizing emotions against neutral. The present findings revealed no effects of either emotion or intensity main for the P1. For both the N170 and FcEP Early, anger evoked a larger amplitude response than happiness for both 20% and 40% intensities. I found no significant main effects of emotion and intensity for the FcEP Late components.

2.4.1 Lateralization Effects on the P1

I did not find any main effects of emotion and intensity for the P1. The P1 is traditionally thought to respond to low-level characteristics of stimuli, such as contrast, luminance, or salience (Batty & Taylor, 2003). Nevertheless, I found a lateralization effect of the P1 such that the amplitude is larger at the right posterior electrode cluster than the left. Even though not always investigated, the P1 seems to arguably react beyond simply low-level visual characteristics. Lateralization of emotion recognition was also reported in studies. For instance, Sedda et al.’s (2013) study showed that patients with the right temporal lobe epilepsy (TLE) were impaired in recognizing negative emotions while patients with the left TLE were intact with emotional recognition. A similar study (Rice, Caswell, Moore, Hoffman, & Ralph, 2018) also found the left TLE patients were less accurate in recognizing all expressions except sadness, and the right TLE patients showed reduced accuracy in all expressions except anger. The present findings are broadly consistent with the suggestion that the initiation of emotional recognition starts as early as the P1 time window.
Studies investigating the effect of emotions on P1 amplitude have failed to produce consistent results. A growing body of evidence suggested the P1 component engages in personal affective judgments of faces (Pizzagalli et al., 1999) and facial expressions where fearful faces increases P1 amplitude compared to neutral faces (Pourtois, Grandjean, Sander, & Vuilleumier, 2004). In the current study, I did not use fearful faces, which may also explain why there is no emotional effect for the P1. Additionally, the absence of emotional or intensity effects on P1 amplitude was also reported by multiple studies (Fruhholz, Jellinghaus, & Jerrmann, 2011; Krolak-Salmon et al., 2001; Sprengelmeyer & Jentzsch, 2006). Study-specific conditions, including a brief stimulus presentation time and the use of low intensity emotions may contribute to the elimination of the P1 effect in this study.

2.4.2 Emotional Effects on the N170

I found an emotional modulation of the N170 amplitude. The current data revealed that angry facial expressions elicited a larger N170 at posterior sites than happy or neutral expressions. This result challenges the typical interpretation of the N170, in which it is purportedly sensitive just to the structural coding of faces (Eimer, 2000). I did not observe significant lateralization of the N170. Although right-lateralization of the N170 has been observed in many studies (Kanwisher, 2006), findings in this study, however, were in line with those of Mercure, Kadosh and Johnson (2011), in that lateralization was absent. Maurer, Rossion and McCandliss (2008) argued that habituation may eliminate the right-lateralization of the N170. In the current study, stimulus images were repeatedly presented, which could explain the absence of the lateralization effect of the N170. Hinojosa et al. (2015) also highlighted that a right-hemisphere dominance in the N170 evoked by emotional expressions was not consistently observed.

Even though I did not have a clear hypothesis concerning the impacts of specific emotions on the N170, I found that N170 amplitude was modulated differentially by different expressions:
anger elicited a greater N170 amplitude modulation than happiness. In addition, N170 amplitude also differed significantly between 40% happiness and 0% (neutral).

Previous studies reporting the response of the N170 to emotions have focused on comparing emotional and neutral expressions. Several studies (Miyoshi, Katayama, & Morotomi, 2004; Wright & Kuhn, 2017) have used two-phase transitional stimuli in which the alternation between stimuli has no inter-stimulus interval, and thus appears as apparent motion. In the study by Wright & Kuhn (2017), an emotional face was preceded by either a neutral or emotional face of the same or different identity. When the stimulus depicted a transition from neutral to either happiness, anger, or fear the amplitude of N170 was enhanced. Similarly, when the transition was from emotional to neutral, the N170 decreased in amplitude. Other studies (Blau, Maurer, Tottenham, & McCandliss, 2007; Stekelenburg & Gelder, 2004) have also observed an enhanced amplitude of the N170 for emotional expressions, especially negative or threatening expressions such as anger and fear. Wright and Kuhn (2017) suggested that emotional expressions consume more processing resources compared to non-emotional expressions. When neutral expressions transit to emotional ones, the consumption of these resources increases and results in an increased N170 amplitude. Behavioral studies (Bar-Haim, Lamy, Pergamin, Bakermans-Kranenburg, & Van Ljzendoorn, 2007; Fox, 2002) found that people are more ready to summon attentional resources when encountering negative or threatening faces. Besides, faces with happy or angry expressions extinguished less readily compared to neutral expressions in brain-damaged patients who had chronic right hemisphere damage and left visual neglect (Vuilleumier & Schwartz, 2001a), which suggests a more sustained response to emotional expressions. Thus, the increased responses to emotional expressions to neutral ones may indicate a prioritized processing of emotional stimuli. Further, emotional information is more salient when stimuli occur concurrently and compete for attention (Kastner & Ungerleider, 2000). In the current study, I found that the
N170 response differs significantly between 40% happiness and neutral (0% expression). This is partially in line with the abovementioned studies even though I found no N170 amplitude difference between anger (both 20% and 40%) and neutral (0%), which might be related to brief stimulus presentation and lower expressional intensity.

Studies investigating the N170 responses between different emotional expressions are far from clear. Batty and Taylor (2003) found fearful expressions evoked a larger N170 amplitude than neutral, happiness, disgust, surprise, sadness, and anger. In another study, fear elicited a higher N170 amplitude than disgust (Almeida et al., 2016). The findings revealed N170 amplitude differences between anger and happiness. Taken together, the N170 data seem to suggest that discrimination of emotional and neutral faces, as well as between emotional faces, seems to occur as early as the first 100-200ms.

Besides, the current findings deviate to some extent from Bruce and Young’s (1986) parallel processing model. If encoding the facial identity and expression are strictly independent and isolated from each other, the N170 should solely respond to either faces or facial expressions (Hinojosa et al., 2015). Although the current study did not involve an interactional experiment between facial identity and expression, the N170 is widely recognized to be face-specific [for example, studies (Caharel, d'Arripe, Ramon, Jacques, & Rossion, 2009; Jemel, Pisani, Calabria, Crommelinck, & Bruyer, 2003) showed N170 amplitude is attenuated with familiar faces]. The emotional modulation of the N170 in the current study, therefore, may suggest non-parallel processing routes of faces (Atkinson & Adolphs, 2011). Alternative models suggested interactive processes between identity and facial expression. For example, Haxby, Hoffman and Gobbini (2000) proposed a modified model of Bruce and Young’s (1986) arguing face perception functions are the result of coordinated participation of multiple brain regions. The decoding of facial expressions in the STS relies on the process of facial identity in the FFA, enabling the process of different identities with distinctive
expressions (Haxby et al., 2000). In a like manner, Calder and Young (2005) suggested that the perception of facial identity and emotional expression was encapsulated in a single representational framework, which some dimensions code facial expression and identity respectively and other dimensions code both. This is consolidated by computing face recognition using the principal component analysis (PCA) and returning that some principal components simultaneously code both identity and expression (Calder, Burton, Millder, Young, & Akamatsu, 2001).

Aside from that, the finding of a significant interaction between emotion and intensity for the N170 further upholds the interactive process of multiple facial dimensions. It is noteworthy that anger as negative emotional expression and happiness as the positive expression were used as stimuli in the present study. The different responses of the N170 may result from the discrimination between positive and negative expressions which may indicate more categorical differences (positive vs. negative emotions) rather than distinction between discrete expressions (anger vs. happiness) (Leppanen, Kauppinen, Peltola, & Hietanen, 2007). Furthermore, enhancement of N170 amplitudes by emotional expressions compared to neutral implies an early process of facial encoding with both valance (positive vs. negative expressions) and arousal dimensions (arguably intensity).

As a result, N170 responses to multifaceted elements of a face may advocate non-parallel and interactive routes of the process between identity and emotional expressions.

2.4.3 Facial Expression Processing over Frontal regions

I observed increased activation in the FcEP early component (120-160ms) for angry expressions relative to happy expressions. Several studies have also reported an enhanced activity at frontocentral region for negative facial expressions, especially fearful expression (Eimer & Holmes, 2007; Smith, 2011). Smith (2011) proposed that the elevated neural
responses to negative emotional expressions may reflect automatic processing over frontal regions which are likely to originate in prefrontal or anterior cingulate cortical regions (Vuilleumier, Armony, Driver, & Dolan, 2001). It is also important to note that the time window of the FcEP early is the same with the N170, the findings may reflect the positive side of the dipolar activity that generates the N170, since results for both components are similar, although the dipolar equivalent of the N170 is typically reported as the VPP, with a more central scalp distribution.

Eimer and Holmes (2007) concluded based on FcEP data that affective stimuli can be processed pre-attentively and automatically. Nevertheless, findings from brain-imaging studies remain inclusive and conflicting. Vuilleumier and colleagues (2001) reported that activation of amygdala by fearful expressions was unassociated with spatial attention. Studies (Vuilleumier & Schwartz, 2001a; Vuilleumier & Schwartz, 2001b) on patients with neglect and extinction of visuospatial processing demonstrated that emotional stimuli can be detected and processed in the absence of attention. On the other hand, Pessoa, McKenna, Gutierrez and Ungerleider’s (2002) study revealed that the amygdala increases in response to fearful faces, but only when sufficient attention was devoted to the faces. Their finding suggests that attention gates the processing of emotional expression in visual cortex.

The amplitude differences between emotions for the later FcEP components (which I termed FcEP late I and FcEP late II) did not reach statistical significance in this study. This result was inconsistent with Eimer, Holmes and McGlone (2003) who noted a sustained difference in ERP amplitude for emotional faces starting from approximately 160ms post-stimulus at frontocentral electrodes site. The absence of such an effect in the present study possibly stems from the current paradigm, in which participants were instructed to direct their attention to surprise faces while other expressions (anger and happiness) were task-irrelevant. On the other hand, the current findings were in agreement with another experiment in the
Eimer et al.’s (2003) study which they found that there were no ERP differences between emotional and neutral faces when participants had to determine whether the lengths of parallel lines were identical between two faces. That is, the faces themselves were task-irrelevant and presumably did not command the same attentional resources. Therefore, data in this study might constitute some evidence that the allocation of attentional resources is required to sustain emotional representation later in processing. Further, my results showed that anger evoked a larger frontocentral activity than happiness, although the difference did not reach statistical significance. If this difference turns out to be real, it may indicate that emotions were processed categorically rather than as one collective response. That said, perhaps the most plausible explanation for the absence of the emotional effect for the late FcEP components is that these components may require stronger and less ambiguous emotional stimuli. This requires further investigation to determine the conditions required to sustain the FcEP response.

2.4.4 Threshold of Identifying an Expression

The current analysis of the N170 data revealed a significant interaction between emotion and intensity. Anger and happiness started differing for the N170 at 20% intensity level. This finding is inconsistent with data reported by Recio, Schacht and Sommer (2014) who failed to find an effect of intensity at the N170 when they compared 60%, 80% and 100% expressions in their first experiment. In their second experiment, they used only 80% and full-intensity expressions. Their results showed that only 80% disgust elicited a larger P1 compared to neutral, and intensity differences did not significantly influence N170 amplitude. Moreover, intensity effects appeared about 200ms onwards and for full-intensity expressions only (Recio et al., 2014). For a single emotion alone, the current result showed that it requires at least 40% for happiness and higher intensity for anger to be differentiated from neutral for the N170. The finding supports previous studies in concluding that happiness is recognized
rapidly and accurately (Palermo & Coltheart, 2004). Recio and colleagues (2014) also observed lower hit rates for anger, especially at low intensity and concluded that anger, as compared to other expressions, benefits from higher intensity for more accurate recognition. The inconsistent findings between this study and Recio et al.’s (2014) study may arise from the complexity of stimuli used in their study. They used dynamic expressions where the contrast between neutral and expressions started at 33ms, so participants had extremely brief time to process at lower intensities (between 0% to the target intensity) since the transition between 0% to the target intensity was completed in 200ms. Likewise, Sprengelmeyer and Jentzsch (2006) found intensity effects on the N170 over parieto-occipital sites, such that increased emotional intensity corresponded to an increased N170 amplitude, regardless of the specific expression. Moreover, the higher intensity level required for discriminating within-emotion than between-emotions could also be ascribed to visually more distinctive characteristics between emotions than between emotional expressions and neutral.

The significant interaction between emotion and intensity may suggest that two different processes occur roughly in parallel when differentiating emotional expressions. That is to say, when differentiation occurs between expressions, it only demands a small difference in intensity, e.g., 20% in my study. This is in line with the idea that the recognition of different expressions can rely on the emotional feature contrast (Calvo, Fernandez-Martin, & Nummenmaa, 2012). On the other hand, when one is required to identify a single expression, since the contrast between emotional features is absent or minimal, it will then demand a larger portion of intensities and the portion required differs between expressions. This idea can be supported by Calvo, Avero, Fernandez-Martin and Recio’s (2016) study, in which happiness has the lowest threshold and can be identified at 20%, followed by sadness, anger, surprise, and disgust at 40% and fear at minimum 50%. Though the threshold I identified for happiness was 40% and anger was higher than 40%, both studies demonstrated that the
thresholds for emotional recognition can vary. The differences in the required intensity level in this study compared to Calvo et al.’s (2016) may stem from disparities in stimuli and methodologies, such as a brief stimulus presentation.

Studies investigating different intensities of facial expression have produced inconclusive ERP findings. The effect of intensity is not consistently present, and the intensity levels required to find effects are variable. In this study, since the P100 was not modulated by emotions or intensities, I have no evidence that the intensity effect was a result of low-level visual differences such as differences in contrast or spatial frequency between images. Therefore, I tentatively assume that there may be a threshold for facial expression intensity detection, but the threshold may vary depending on how much attention was devoted and perhaps on other idiosyncratic aspects of the experimental tasks. In previous studies, participants were either instructed to identify emotions (Leppanen et al., 2007; Recio et al., 2014) or respond to gender (Sprengelmeyer & Jentzsch, 2006), whereas in the current study, they were required to pay attention to surprise faces, which only accounted for 10% of the total trials. In other words, their attention was not being interfered or directed away while processing the stimuli, which may explain the presence of intensity effect in this study even at lower intensity levels.

Based on the abovementioned ERP studies on emotional intensity, the processing of intensity information seems to be ambiguous, especially for lower intensities at early stages. The interpretation of early ERPs in facial expression processing is complex because multiple dimensions of information are simultaneously present in the image of a face.
2.4.5 Facial Expression Processing

Luo, Feng, He, Wang and Luo (2010) proposed three stages of facial expression processing. In the first stage, they claimed that facial expressions – especially threatening ones like anger and fear – are distinguished. The temporal window for the first stage to occur is approximately coincident with the P1 component (Zhang, Luo, & Luo, 2013). In this study, P1 amplitude did not differ between anger and happiness. Possible reasons may be the lower stimulus intensities and brief presentation times in my experiment. Facial expressions with lower intensity increase the difficulty in differentiating emotions, which might interrupt the automatic processing at this early stage of processing. In the second stage, following the rapid detection of particularly threatening expressions, emotional faces are processed as a whole and are distinguished from neutral faces. This occurs around the time of the N170. My findings, however, did not fully support the stage-like model described by Luo et al. (2010), because I found that anger evoked a larger N170 than happiness even at 20% intensity level. Though I found anger (both 20% and 40%) did not differ from neutral (0%), happiness differs significantly from neutral at 40% intensity level. The result demonstrated that the brain is able to differentiate different emotional expressions during the N170 window. In the third stage, different expressions can be defined and distinguished in a broader time window, beginning at around 300 ms. In this study, expression-related differences in amplitude of the later FcEP components failed to reach the statistical significance threshold. The differentiation between expressions was processed in the stage 2 window (the N170 window) in temporal order in the current study but, for some reason, this effect did not sustain into later time periods. This may be due to the low intensities of the expressions, and perhaps to the task-irrelevant nature of these expressions. It would be interesting to see whether later, sustained effects were obtained for high-intensity, but still task-irrelevant threatening expressions.
I inclined to assume that the facial expression processing may not occur as a linear, stage-by-stage process as proposed by Luo et al. (2010), but perhaps occurs in a more flexible, cascade-like process depending on the specific task and situation. The three-stage processing of facial expression is probably over-simplified for several reasons. First, several lines of evidence (Batty & Taylor, 2003; Fruhholz et al., 2011; Sprengelmeyer & Jentzsch, 2006) reported global processing of emotional stimuli during the P1 window. Enhanced P1 amplitudes were reported for threatening (e.g., fearful faces), pleasant (Brosch, Sander, Pourtois, & Scherer, 2008; Utama, Takemoto, Koike, & Nakamura, 2009) and even disgust stimuli (Utama et al., 2009). This could imply that the differentiation between expressions may start as early as the P1 but may just be at a “preparation stage” and a global level.

Second, studies (Chen et al., 2015; Rossignol et al., 2012) including the current one, showed that the N170 is sensitive not only between emotional and neutral faces, but also between different expressions. Further, the meta-analysis by Hinojosa et al. (2015) revealed that anger increased N170 amplitude the most, followed by fear and happiness. They found that sadness and disgust also enhanced N170 amplitude but failed to reach statistical significance. Thus, they contended that not all facial expressions are equally sensitive to the N170, but that there is a hierarchical response to different expressions – particularly negative ones. Compared to other expressions, fear and anger require faster responding and reaction (Hinojosa et al., 2015). Additionally, the N170 is reported to be sensitive to race (Tortosa et al., 2013), attraction (Halit, Hann, & Johnson, 2000) and aging (Wiese, Schweinberger, & Hansen, 2008). Taken together, evidence suggests that N170 is influenced by multiple dimensions of faces and can differentiate beyond merely between emotional and neutral faces. Third, Luo et al. (2010) did not take other dimensions of expressions, such as variation in intensity, into account. Different intensities of emotional expressions may advance or postpone the process of emotional stimuli.
Altogether, I would assume that the whole process of facial expressions in Luo et al.’s (2010) three-stage processing can in appropriate circumstances actually be condensed and accomplished around N170 time window. At the P1, attention or other cognitive resources are ready to be consumed but possibly expended all at once, since drawing a large pool of attentional resources at such an early phase can burden the system. These resources, therefore, can be reserved and are sufficient in supply for fully processing facial expressions at a later stage. At the N170, cognitive resources can be allocated to face processing with multiple dimensions, including emotional expressions, intensities, age, attraction, etc. exerting their influence. More fine-grained and complicated processes (e.g., further discrimination between expressions with intensities, physical responses, etc.) will then take place beyond the N170 time window. These assumptions can be tested in future studies. A further thought is that emotional process may not occur in a strictly time-locked fashion, but may occur a different times depending on the situation. Studies report inconsistent time windows for different “stages”; for instance, I found discrimination between emotions occurs earlier than Luo et al. (2010) report. The discrepancy can be explained if processing emotional expressions is not strictly time-locked, but can vary according to different factors such as intensity, attraction, or familiarity. Again, we need further studies to test this idea.

2.4.5 Limitation and Future Direction

Although findings reported here support the notion that emotion and intensity in facial expressions are processed at a fairly early stage, it is also worthwhile to elaborate on some issues that may limit my interpretations. First, it is important to point out that my stimuli were limited to only happiness and anger, which also represent positive and negative categories. One might argue that the current findings better reflect valence differences between expressions rather than discrete facial expression. Indeed, I am unable to completely rule out such a possibility. However, it is evident that neutral expressions (0% in this study) did not
differ from emotional expressions as there was no statistical significance between 0% and other intensity levels. Therefore, it is plausible to assume that the distinction between emotions can be processed at the N170. Future studies can incorporate facial expressions that are less valence-distinctive, such as anger, fear, and sadness, to better clarify ERP differences and test the generalization of my results to other emotions. Second, I only included low intensities (20% and 40%) in this study as I was interested in subtle expressions, and if a threshold of emotional expressions recognition was present. Of course, it is arguable that the 20% gap between intensities in this study was probably too wide to identify a threshold. Hence, further studies can investigate intensity levels from 40% above to locate the threshold or within 40% to consolidate or disprove the present findings. Third, I used static images and morphs as stimuli to present emotional expressions and intensities. These are widely used in similar research but have limited ecological validity. While morphing allows good control and quantitative variation between intensities, it assumes a linear unfolding of expressions in spatial and time course, which may be too simplistic. Thus, future studies can incorporate stimuli that can precisely control the intensity levels and reflect the dynamic nature of facial expressions, for example, with the use of computerized avatars. Fourth, the sample size in this study could be larger to confirm the effects, although a sample size around 20 participants is common for an EEG/ERP study like this one (Achaibou, Pourtois, Schwartz, & Vuilleumier, 2008; deSilva et al., 2016).

To summarize, I found evidence that supported the early processing of facial expressions and sensitivity to emotional intensity. Anger, compared to happiness, elicited a larger N170, which may be indicative of a more interactive processing of a face. This means structural encoding, emotional expressions, and intensities can be processed in parallel if not interactively. Future research should incorporate more basic emotions to minimize the categorical distinction and generalize the findings. The interaction between emotion and
intensity showed that a 20% intensity level was demanded to discriminate between emotions, while a 40% or higher intensity level was required to discriminate within an emotion. The current findings also suggested a 40% intensity level may be a borderline to fully distinguish emotions, and future studies can incorporate intensity levels above and below 40% to assess more directly the mechanisms of intensity processing in emotions. Lastly, the absence of differences between emotions or intensities at frontocentral region may be inferable to insufficient attention devoted to the task. More studies can manipulate attention level to investigate if attention is required in processing emotions at frontocentral area.
Chapter 3 The Uncanny Valley

3.1 The Emerging of Virtual Stimuli in Facial Expression Study

We see faces all the time. We broadcast our feelings via facial expressions of emotion and understand others by recognizing them. In Chapter 1, I have reviewed theories and studies of emotion and expressions. Much of what we know about faces and facial expressions derives from studies of images, for instance, Picture of Facial Affect (POFA) (Ekman & Friesen, 1976) and The Extended Cohn-Kanade Dataset (CK+) (Lucey et al., 2010). Despite the importance of the stimuli in facial expression study, to our surprise, there is no systematic review in the literature regarding either the quality of the facial stimuli or facial stimuli control. The core problem of face stimuli is probably derived from “the tradeoff between ecological validity and control of experimental conditions and extraneous variables” (Ferreira-Santos, 2015, p. 236).

Simple schematic faces represent an abstract form of real faces (Ohman, Lundqvist, & Esteves, 2001) (see Figure 3.1). An advantage of using schematic faces in studies is the precise control of stimuli. Schematic faces can be controlled for identical physical features, or emotional expressions can be equally discriminable from neutral (Purcell & Stewart, 2010; Ohman et al., 2001). Additionally, researchers are able to use schematic face stimuli to create exaggerated or caricature expressions (Makarainen et al., 2014). Furthermore, the use of schematic faces as stimuli is demonstrated to generate comparable results with photo face stimuli. For example, Wehrle, Kaiser, Schmidt and Scherer (2000) demonstrated that emotion recognition of synthetic images of facial expressions received comparable recognition rate to posed photos even though the images were produced by schematic lines. Maratos, Garner, Hogan and Karl (2015) also observed similar N170 response of schematic faces with real faces.
Morphs and videos as ways of retaining or regaining some stimulus control have their limits too. Morphing is a technique that transits one image into another seamlessly (Steyvers, 1999). This technique is also frequently used to transform one expression, usually neutral, to another expression to create different intensities of expressions, or combine expressions (e.g., happily disgust). However, stimuli are still far from being perceptually “real” and present some unique challenges. The mechanism by which morphing may not represent the nuance that is available in facial expressions. In other words, it does not exactly imitate how human faces produce different levels of expressions. Typically, with two faces, one represents a neutral face, and the other one displays a full expression. Morphing software is then used to merge the two pictures and create facial expressions at various intensity levels (e.g., 20%, 40% …). This technique enables researchers to manipulate facial expressions precisely, yet it does not characterize the way people generate facial expressions as we might do in real-life. To put in a simpler way, people do not portray a smile by merging a neutral expression and 100% happiness in the real world.

Short video clips of actors/actresses posing facial expressions are also widely used as dynamic stimuli in facial expression studies. Compared to static images, dynamic faces were rated as more intense (Rymarczyk, Biele, Grabowska, & Majczynski, 2011) and elicited stronger and spontaneous facial muscle mimicry (Sato & Yoshikawa, 2007). However, video
clips open up another set of problems. It is difficult to control actors/actresses’ way of posing expressions (e.g., people have different ways to express smile or happiness) and their physical appearance (Joyal, Jacob, Cigna, Guay, & Renaud, 2014); many face stimuli lack of authenticity by appearing posed or “staged”. Moreover, dynamic video stimuli can introduce uncontrolled variance, like various time periods of stimuli presentation (Ferreira-Santos, 2015) since there is no evidence suggests that different expressions are generated by the same amount of time.

Computerized avatars therefore provide an elegant way to reconcile the precise control over face stimuli and maintain a certain level of ecological validity at the same time. The advancement of computer technology has also popularized the use of virtual agents in facial expression research. FaceGen, for example, is a software program developed by Roesch et al. (2011) aiming to create realistic 3D facial avatars based on the Facial Action Coding System (FACS) (Ekman et al., 2002). An example display look of FaceGen is shown in Figure 3.2. The software allows experimenters to have systematic manipulation of multiple dimensions in a face, including expressions, age, ethnicity, and gender (Roesch et al., 2011). It offers accurate control over facial action units and have precise control of the stimuli. Compared to Ohman et al.’s (2001) schematic faces, FaceGen has increased ecological validity but still looks unreal. In comparison, the AFS is more sophisticated by incorporating the FACS and virtual nervous system (the Brain Language) that embody real-time learning and sensorimotor interaction to produce hyper-humanlike face models (Sagar et al., 2015). Experimenters can generate even subtle changes in facial expressions by meticulously adjusting the parameters of each AU, which is another advancement compared to FaceGen.
With the arrival of highly human-like avatars, studies have reported an equal or better recognition rate of facial expressions from virtual avatars except disgust as compared to human faces (Dyck et al., 2008). In brief, the use of virtual faces evokes facial mimicry of similar facial muscle groups and generate similar recognition rate as photo faces (Joyal et al., 2014).

Efforts to make synthetic faces/expressions span a large range of realism, from schematic lines to state-of-the-art computation (the AFS). It is worth noting that there may be a need to distinguish efforts to many synthetic faces for research purposes (e.g., Ohman, or FaceGen) and those that are more motivated by other demands (e.g., movies or animation). Although hyper-realism is probably the ultimate goal of synthetic human agent designers, increments in realism beyond a point can cause a paradoxical emotional rejection or repulsion in observers – a phenomenon that has been termed the “uncanny valley”.

### 3.1.1 The Uncanny Valley in Robotic/Simulated Faces

New technologies have facilitated in creating increasingly realistic and human-like computer-generated characters, especially in video games, movies, or human-machine interaction areas.
Many highly realistic animations with non-human subjects in movies like *Toy Story*, *A Bug’s Life*, and *Finding Nemo*, are well-liked and popular. Yet another favorite animation known as “The Adventures of Tintin” has spurred discussion of an “eeriness” feeling in response to the character Tintin, which was described in *Vulture* magazine thusly: “Tintin looks simultaneously too-human and not human at all, his face weirdly fetal, his eyes glassy and vacant instead of bursting with animated life” (Buchanan, 2011).

Although some animations like *Finding Nemo* are deliberately made less realistic by animation companies like Disney or Pixar than they were able to do technically to ostensibly retain the sense of animated fun rather than hyper-realmism, it is a general goal of many animators and digital artists to create highly realistic digital representations of faces and expressions (Cosker, Eisert, & Helzle, 2015). They believed that highly realistic characters enhance the users’ experience. Unfortunately, the positive feeling holds until a certain point where the character is nearly– but not yet human. At some point, a sense of discomfort and weirdness, known as the “uncanny valley” experience, emerges (see Figure 3.3). However, as the humanness keeps increasing, and goes beyond the point, the weirdness sense decreases and the positive feeling returns. The whole process is termed the “uncanny valley cliff”.

The term “uncanny” was firstly introduced by Jentsch (1906) from the German word “unheimlich”, which simply means when someone is not “at ease” or facing something or situations that are foreign. He further used the term to describe a situation where people can no longer differentiate what is real or not real. However, he did not clarify the concept further, and, instead, he argued that a definition is not as crucial as examining the affective excitement of being uncanny. Jentsch (1906) further claimed that if one feels nothing is uncanny, he or she might be in question of psyches as he described being uncanny as a fundamental condition, though he claimed with the absence of elucidating what being uncanny actually refers to. In Jentsch’s (1906) work, he proposed an early insight of uncanny
and attributed such feeling to psychical conflicts derived from uncertain things or circumstances.

More than half a century later, Masahiro Mori (1970) advanced a similar concept and termed it “bukimi no tani” which is “uncanny valley” in Japanese. In Mori’s paper, he plotted the uncanny valley on two dimensions, namely human likeness, and shiwa-kan which was translated into familiarity, affinity, or emotional engagement (see Figure 3.3, below). Besides, Mori (1970) also emphasized that the motion of robots tends to amplify the peaks and valleys.

Mori (1970) did give a sense of uncanny valley phenomenon with examples. However, a single example (or few examples) certainly bears the problem of lacking generalization. It failed to take individual differences into account as the two dimensions – human likeness and affinity are fundamentally subjective. Furthermore, if considering the definition of uncanny valley, the sense of eeriness is caused by things not completely human, like a zombie or corpse (Katsyri, Forger, Makarainen, & Takala, 2015). Hence, examples described by Mori (1970) cannot fully explain why people have discomfort towards those objects. Humans disengage with a corpse not simply because it is not human anymore, but because it is morbid in nature. The same thing goes with a zombie. Although its existence is entirely debatable, people are afraid as it is life-threatening – and horrific in its very appearance. For example, when a person sees another individual with an injury to his or her face or even a birth defect (such as a cleft lip not corrected by surgery) the person might avoid looking at the individual because it creates a feeling of discomfort. Such feeling is probably not derived from lack of humanness. Consequently, zombies or corpses are not able to accurately represent the uncanny feelings suggested in the theory.
Mori (1970) theorized that when people are exposed to robotic characters that are uncommon in real life circumstances, they will feel disturbed. The uncanny feeling could also occur when people perceive a highly human-like robot if they realize that it is not human (Seyama & Nagayama, 2007). Furthermore, Blow, Dautenhahn, Appleby, Nehaniv and Lee (2006) argued that with the increased familiarity of robots, the feeling of eeriness would be reduced since people will have become more acclimatized to them. As a result, if familiarity interacts with habituation, whether it will skew or distort the uncanny curve remains unclear.

Despite the fact that today the term uncanny valley is widely used in popular media to discuss animated films, the concept is pretty much under development. In the following sessions, I will delineate three key elements of Mori’s theory of the uncanny valley, which are human likeness, affinity, and effect of movement.

![Figure 3.3 Illustration of Mori’s uncanny valley curve showing a non-linear relationship between human likeness and affinity. Mori hypothesized that movement will amplify the uncanny valley. Adapted from Mori (1970) translated by MacDorman and Kageki (2012).]
3.1.1.1 Human-likeness

In Mori’s original paper, he did not further specify the concept of human likeness, nor did he describe how it could be measured or quantified. Instead, he visualized different levels of human likeness with examples.

He used an industrial robot that usually has only an extended mechanic arm to represent the least human-like characters, which bear no similarity to humans. He emphasized that it is the function of industrial robots that matters rather than the appearance or whether they look human or not. Hence, industrial robots sit near “0” of Figure 3.3, where the least human likeness and least affinity are experienced. For the next level of human likeness, Mori represented this with a toy robot and stuffed animal to exemplify moving and still status respectively. The appearance of both the toy robot and the stuffed animal starts resembling external human features such as a face, a torso, and four limbs. These toys are popular among children, so they are rated on or near the first peak in the curve (see Figure 3.3). For the valley, which Mori exemplified with a corpse or zombie to correspond to static and moving status, both the corpse and zombie have full or the majority of features of a human, but people react to them negatively, partly due to their creepiness and surviving nature of danger avoidance. Lastly, the curve terminates with a normal human.

What we can take from Mori’s examples is that some stimuli may have fully or nearly human features, but they still lack something, like Tintin (Buchanan, 2011). Maybe there are some subtle clues of humanness that make us uncomfortable. This dimension of the uncanny valley has therefore been the subject of much concern and speculation amongst both animators and scientist, alike.
3.1.1.2 Shinwa-kan – Affinity

The term “shinwa-kan” in Japanese has no direct translation into English. A common translation is familiarity or likability because “shinwa-kan” “is not a commonly used word, nor does it have a direct equivalent in English” (Bartneck, Kanda, Ishiguro, & Hagita, 2007, p. 369). In a later translation by MacDorman and Kageki (2012), “shinwa-kan” was translated as affinity. Bartneck et al. (2007) questioned the translation and highlighted that familiarity depends on the amount of exposure to robots. If people experienced robots more, the familiarity increases, and the associated uncanny feeling would diminish. For that reason, it may imply a short duration of uncanny valley, which could contradict to the real-life situation. Bartneck et al. (2007) further detailed that familiarity is unlikely to decrease with the increasing exposure of robots, whereas people will not automatically like all robots regardless of the exposure frequency. As such, they proposed that the term “affinity” or “likeability” appears to be more appropriate to conceptualize this dimension.

Owing to the fact that the affinity dimension was also underdeveloped in Mori’s original paper, and the ambiguous nature of the term, Bartneck, Croft and Kulic (2009) developed what they coined “The Godspeed Questionnaire” in order to provide a standardized measurement for human-robot interaction. The questionnaire covered five factors, including “anthropomorphism”, “animacy”, “likability”, “perceived intelligence”, and “perceived safety”, and applied a 5-point scale of semantic differential items for each concept. In semantic differential scales, participants are required to rate on a continuum with two bipolar words; for instance, fake – natural.

Although the questionnaire attempted to conceptualize the factors, Ho and MacDorman (2010) pointed out that semantic differential items in the Godspeed questionnaire were not empirically tested for reliability and validity. Some concepts like “perceived intelligence” in Bartneck et al.’s (2009) questionnaire were less pertinent to the idea of the uncanny valley.
Further, the items in the questionnaire were significantly correlated, such as “anthropomorphism”, “animacy”, “likeability” and “perceived intelligence”, which indicates these items may measure very similar concepts. Hence, Ho and MacDorman (2010) proposed alternative concepts to assess people’s perception of anthropomorphic objects, namely “perceived humanness”, “eeriness” and, “attractiveness”. These alternative indices were not significantly correlated and were validated. In their proposed questionnaire, designed to replace the Godspeed instrument, there are 19 semantic differential items grouped into three concepts. Five items were included in the “attractiveness” index, while eight and six items were included for the “eeriness” and “humanness” indices, respectively. Nevertheless, Katsyri et al. (2015) also raised three potential questions of Ho and MacDorman’s (2010) questionnaire. First, like the Godspeed Questionnaire, Ho and MacDorman (2010) also utilized semantic differential items under each concept. The problem is that some items caused confusion or did not necessarily measure the particular concept. For instance, including “spiteful – well-intentioned” as a measure of warmth is confusing; the same thing can be applied to “ordinary – supernatural” under humanness. Secondly, eeriness projects to emotional responses. Katsyri et al. (2015) argued that items under eeriness in the questionnaire use atypical words to describe emotions, which might not measure what it was intended to measure, hence it may also have a validity issue. Examples like “ordinary – supernatural” or “uninspiring – spine-tingling” are arguably not entirely associated with eeriness. Finally, the concept of “familiarity” was left out in Ho and MacDorman’s (2010) questionnaire. Familiarity is an earlier and common translation of “shinwa-kan”. The absence of the term failed to mirror the affinity dimension of Mori’s uncanny valley theory. As a result, Katsyri et al. (2015) concluded that for future scales, the wording could be simplified to a common metric to reflect affinity.
In addition, Katsyr et al. (2015) also provided some insights of understanding affinity as they believed this dimension is associated with perceptual familiarity and emotional valence. According to them, perceptual familiarity refers to identifying and recognizing similar features of one object to the object that an observer is familiar with. Emotional valence encompasses positive and/or negative emotions of an observer projected to the perceived features. Taken together, the particular dimension has encapsulated rather broad sub-concepts.

3.1.1.3 The Effect of Movement on the Uncanny Curve

From animals to humans and robots, movement is an integral element in their representation. Mori (1970) suggested that movement magnified the entire uncanny valley curve by increasing the peak and deepening the valley. He proposed that when an industrial robot moves like a human, people begin to accept it as more human. In the same vein, if a prosthetic hand starts moving, the uncanny feeling amplifies.

Saygin, Chaminade, Ishiguro, Driver and Frith (2012) used fMRI repetition suppression to study how human action perception system (APS) responds to a human agent regarding the appearance and motion. They had participants watch video clips of three characters: a human being with biological appearance and motion, a robot with mechanical appearance and motion, and an android with biological appearance and mechanical motion. They found suppression effects were more prominent for the android in contrast to the human and robot in anterior intraparietal sulcus bilaterally. We as human are experienced in associating human appearance with biological movements and machine appearance with mechanical movements. The incongruence of the android, which is characterized by human appearance and mechanical movements, has therefore flagged a larger prediction error. Saygin et al. (2012) concluded that the brain will experience more prediction errors when inspecting a character with a human appearance but does not move like a human. Such errors reflected by
the brain activities may explain the discomfort feelings in seeing the android, therefore may explain the uncanny valley concept. Another study (McDonnell, Breidt, & Bülthoff, 2012) also asserted that movement influences how familiar or appealing we think a synthetic avatar is. An unappealing avatar was rated more so when adding motions.

However, not all studies agree with the effect of movement. For instance, Piwek, McKay and Pollick (2014) used full-body computer-generated characters with a range of controlled qualities of motions of joints, from natural movements to distorted movements. They reported that increasing qualities of motion increased the acceptability rating of the character.

Similarly, Thompson, Trafton and McKnight (2011) incorporated computer-generated avatars (a Whole-body male and a Mannequin) with the manipulation of three kinematic parameters of gait. Participants were required to watch a 4s video of one of the avatars and rate for humanness, familiarity, and eeriness. Their results indicated no evidence supporting Mori’s assumption of movement. They demonstrated that when avatars move more like a human, observers tend to perceive them more human and project more positive emotion and less strangeness.

Altogether, the studies regarding movement remain inconclusive. Considering that motions involve different kinematic gestures, smoothness levels, and amplitudes of movement, combinations of these variables could affect viewers’ overall perceptual experience and therefore affect the result of experiencing the uncanny valley. Thus, the complexity of the question maintains the need to assess movement as a critical variable in the uncanny valley phenomenon.

3.1.2 Explanations of Uncanny Valley

Although Mori described the uncanny valley in a simple and straightforward manner, the phenomenon appears too complicated to be explained by a single factor. Predominantly, four
explanations of the uncanny valley are particularly discussed in the literature, which are evolutionary hypothesis, cognitive dissonance, categorization ambiguity, and perceptual mismatch. It is also worth noting that these explanations are neither mutually exclusive nor sufficient to explain the uncanny valley phenomenon as stand-alone theories.

3.1.2.1 Evolutionary Hypothesis

The uncanny valley in Mori’s work was illustrated by corpses and zombies, which are dead (or presumed dead) in nature. Animals, including primates, will hide, bury, or escape from the dead ones to avoid dangers and threats (Spennemann, 2007). As a consequence, when seeing a human agent with a profoundly human appearance that appears similar to a non-living creature, our surviving and defending mechanism will be activated. Moosa and Ud-Dean (2010) posed that dead objects elicited a fear response because of their equality to danger. Moving human agents are “creepier” than still ones as they represent a higher level of danger. Although danger avoidance successfully explained why people may dislike corpses, it does not adequately explain the uncanny valley in the context of incorporating virtual agents in real life, e.g., why people emotionally reject some 3D animated characters like Tintin, which appears to be obscure if explained by danger avoidance.

MacDorman and Ishiguro (2006) brought up the assumption that human androids fall into the uncanny valley because they were processed as humans by our brains, but are not really human at all. Human beings have neuronal origins to process faces and other human features automatically and functionally. When a human android looks human, our brain assesses and matches every possible feature of the android. If the android fails the inspection, which violated our expectation for it as a human, a sense of discomfort emerges. Steckenfinger and Ghazanfar (2009) advocated that to study the uncanny valley via an evolutionary perspective, response of nonhuman species should be examined so that we can assess whether the uncanny valley is human-specific. They employed non-human primates – monkeys and had
the monkeys view three render-types of monkey faces: unrealistic synthetic faces, realistic synthetic faces, and real monkey faces. Their results elucidated that monkeys preferred looking at the real faces and unrealistic synthetic faces as opposed to realistic synthetic faces. They interpreted these findings as a demonstration of the uncanny valley effect for the non-human primates, which provides evidence for the evolutionary root of these processes, as monkeys are very unlikely to have “learned” to process avatars (Steckenfinger & Ghazanfar, 2009).

Lewkowicz and Ghazanfar (2012) asserted that the uncanny valley experience has its origins in early, highly selective, and species-specific perceptual experience. Most individuals are extensively exposed to the faces of their own kind and develop efficient and effective face processing skills. In infancy, individuals learn the prototypical face of their own species. Human newborn babies are attracted by face-like patterns even though the patterns are merely geometric shapes resembled a real face (Cassia, Turati, & Simion, 2004; Lewkowicz & Ghazanfar, 2012). At around 6-months old, infants displayed a marginal preference to avatar faces than human faces. However, at 12-month-old infants started exhibiting a preference to human faces than avatar faces. The preference shift supported the early developmental expertise in face perception and the rejection or less acceptability to faces that do not fit the prototype precisely (Lewkowicz & Ghazanfar, 2012). In contrast, Minato, Shimada, Ishiguro and Itakura (2004) found that when using a full-body android avatar found that some infants around 13-months of age were curious and attracted by the android while older children (aged 3-5 years) rejected and were afraid of the android. The inconsistent findings of preferences to human faces vs. avatar faces of infants at around 12-13 months from the two studies (Lewkowicz & Ghazanfar, 2012; Minato et al., 2004) can be attribute to stimuli difference. The stimulus in Minato et al.’s (2004) study was a whole-body android, while Lewkowicz and Ghazanfar (2012) used still images of only faces. Compared to still
images, an android contains more perceptual information and is more complicated to process. Therefore, 13-month-old infants may just be curious about what they see, holistically, as opposed to more specific processing for faces.

The uncanny valley effect appears to be fundamental and is likely to occur in other species (e.g., non-human primates). This could suggest an important role for the uncanny valley effect in social/emotional processing of faces and bodies. Although, it is also worth noting that the evidence for a link between the uncanny valley experience and evolutionary pressure is still weak.

3.1.2.2 Cognitive Dissonance

Festinger (1957) proposed cognitive dissonance theory and argued that people desire consistency and harmony between attitudes and beliefs. Failure to achieve consistency will cause mental discomfort (dissonance), and one will seek to eliminate the dissonance by changing some cognitive aspects. Given that, when people see a highly realistic computer avatar, two contradictory thoughts will be generated: “I know this is not a person since it is a machine”, with a realistic human appearance, “I think this is a person because it looks/acts like one”. Therefore, the more realistic an avatar appears, the more conflicts an observer generates, thus resulted in cognitive dissonance (Tondu & Bardou, 2011). Tondu and Bardou (2011) had then re-interpreted the initial uncanny valley curve and incorporated the idea of cognitive dissonance to explain this phenomenon. They divided the curve into two sides from the “valley” of Mori’s graph, left and right, which represent “machine” and “human being” fields respectively (see Figure 3.4). They further argued that cognitive dissonance could occur on both sides of the valley. On the left-hand side, dissonance derives from the conflict between knowing these are artificial characters and people’s inclination to deliver emotions because of their human forms. On the right-hand side, dissonance comes from the conflict
between accepting the characters as human and the deviation of these characters (e.g., dying or severely handicapped people in Figure 3.4) from a “normal and healthy” human.

Furthermore, Tondu and Bardou (2011) identified two peaks before and after the valley as points where minimum dissonance occurs. The rightmost peak undoubtedly represents a human being or future avatars that appear human without eerie and cold feelings. Bryant (as cited in Tondu & Bardou, 2011) portrayed the first peak as an object with enough humanness to acquire empathy from an observer, but simultaneously not enough humanness to trigger emotional disengagement (e.g., the C3P0 robot from Star Wars in Figure 3.4). This provides the opportunity for further processing and may allow avatars to reach the second peak of the valley where certainty of humanness is more concrete; hence, allowing easy categorization as a real human face.

![Figure 3.4](image)

*Figure 3.4* Demonstration of examples of the two sides of the uncanny valley: the left-hand machine side and the right-hand human side.

### 3.1.2.3 Categorization Ambiguity

Categorical perception was initially found for auditory stimuli in which speakers categorize phonemes to discriminate speech sounds (Liberman, Harris, Hoffman, & Griffith, 1957). In general, it is the tendency to perceive stimuli that vary continuously along some dimensions as belonging to one of several categories rather than as continuously varying. This concept
was then extended to other perceptual stimuli and applied to explain the uncanny valley as well. According to Harnad (1987), a hallmark of categorical perception is that it is easier to discriminate pairs of stimuli straddling the categorical boundary than pairs equally distanced, but that fall into the same category. Categorical perception can occur in many other instances, including phoneme perception between *ba*'s and *pa*'s (Liberman et al., 1957), color perception characterized by different wavelength (Pilling, Wiggett, Ozgen, & Davis, 2003), or facial expressions such as basic emotions (Ekman, 2003).

Cheetham, Suter and Jancke (2011) proposed that the concept of categorical perception is associated with Mori’s dimension of human likeness (DOH). They argued that the DOH could be represented as two categories: human and non-human (or “avatar”). Pairs of faces that fall on either side of the avatar-human categorical boundary can easily be distinguished perceptually.

They employed a continuum consisting of 13 avatar-human morphs as stimuli to test Mori’s DOH. In addition, the morphs were linear with 8.33% increments in physical appearance from 0% human (or 100% avatar) to 100% human (or 0% avatar). Furthermore, morphs with 33% difference along the continuum were used in a subsequent fMRI study. Participants were required to categorize the pictures of either human or avatar rapidly and accurately. They observed that people spent a longer time in categorizing 33% and 66% human likeness morphs compared to 0% and 100% human likeness. fMRI results revealed that the striatum and medial temporal lobe responded differently to the change of directions along DOH indicating categorical difficulty. Cheetham et al. (2011) concluded that human likeness was not reflected by how similar an object is close to human appearance, but rather suggested that the perception of humanness is a categorical process.
Cheetham and colleagues (2011) further argued that when pairs of stimuli sat on the boundary between two categories; the perceptual discrimination is more laborious than for those stimuli distanced from the categorical boundary, thus taking a longer time to respond. In this vein, they re-conceptualized Mori’s theory and claimed emotional disengagement is most likely to occur due to categorical ambiguity when a stimulus is located at or near the categorical boundary. Findings from their fMRI study indicated the avatar and human faces fall under different categorization problems, and they tend to involve different strategies in the processing. Since Mori did not clarify the human likeness concept, the possibility of differences in human likeness within the human category should also be considered.

Nevertheless, Weis and Wiese (2017) pointed out that Cheetham et al. (2011) did not explain what causes the difficulties in categorical discrimination, which Weis and Wiese (2017) suggested that it might be due to cognitive conflict from increasing cognitive load during categorization. They employed a mouse tracing technique to measure cognitive conflicts and a continuum of 21 morphed faces using a robot and a human face with 5% morphing steps as stimuli. Participants were required to categorize each picture into human or non-human. Results indicated that agents with 70% human likeness fell into the uncanny valley, and induced the maximum cognitive conflict. They also discussed that it was unclear whether the induced cognitive conflict was caused by categorical processing or inspection of lower-level cues as perceptual-mismatch hypothesis implies (see section 3.1.2.4, below).

3.1.2.4 Perceptual Mismatch

According to the hypothesis of perceptual mismatch, negative emotion is elicited by an inconsistency between human features and perceived features of artificial characters, which therefore explains or predicts the uncanny valley effect (MacDorman, Green, Ho, & Koch, 2009). For instance, artificial eyes on a human face or vice versa can cause unease feelings of an observer (Katsyri et al., 2015). Similarly, Seyama and Nagayama (2007) morphed human
faces with dolls, masks, and computer-generated characters as stimuli and found that larger-than-normal eyes produced feelings of eeriness in participants. Likewise, MacDorman et al. (2009) also reported that atypical facial proportion, especially eyes-face mismatch of virtual characters increases the uncanny feeling.

Another interpretation of perceptual mismatch suggested the notion of sensitivity to deviation from human features in computer-generated faces (Katsyri et al., 2015; MacDorman et al., 2009). The human visual system excels in perceiving and discriminating artificial faces. Even for highly realistic synthetic faces, untrained observers exhibit no difficulty in distinguishing those faces from real ones (Farid & Bravo, 2012). As a result, they found people’s sensitivity was intensified when confronting highly humanly realistic and photo-realistically textured artificial faces. With such stimuli, the visual system started inspecting possible cues strictly to determine if the synthetic face is human or not.

The abovementioned studies (MacDorman et al., 2009; Seyama & Nagayama, 2007) highlighted the proportion of eyes to the face is the key (or at least a key) to the uncanny feeling. Following this, Mitchell et al. (2011) built upon Tinwell, Grimshaw and Williams’ (2010) study on synchronization between facial features and sound of a virtual character. Tinwell et al. (2010) found a mismatch between lip-movement and sound of the virtual character exaggerated the feeling of eeriness. The more the visual-auditory cues were synchronized, the less strange the character was rated. Likewise, Mitchell et al. (2011) had participants view video clips of a robot and a human character under two conditions: figure-voice match and figure-voice mismatch. The videos were rated on humanness, eeriness, and warmth of the characters. Their results supported Tinwell et al.’s (2010) findings and suggested that two figure-voice mismatch conditions scored the highest on eeriness. Makarainen et al. (2014) also manipulated the magnitude of facial expression and level of realism to examine if exaggerated facial expressions lead to the uncanny valley. They
reported that when the level of realism decreases, the intensity of a facial expression decreases. To increase the emotional intensity, one could exaggerate the expression. However, by doing so, it tends to make the face stranger and creepier, but if the face is less real, the perceived eeriness is less. As exaggerated facial expressions are not typical of human interactions, lower levels of realism usually involve some exaggeration of expressions in virtual characters. As such, these images do not match how our visual system defines humanness and elicit a negative emotional response.

MacDorman and Chattopadhyay (2016) proposed a very similar theory to perceptual mismatch termed realism inconsistency. They predicted that inconsistent features of an anthropomorphic character associated with realism could result in conflict inferences of determining if the character is a real human being. Additionally, due to the inconsistency and mismatch between neuro-cognitive expectation, increased prediction error, and more negative emotional responses are produced (Cheetham et al., 2011; Saygin et al., 2012). In the study, they manipulated two sets of features: 1) eyes, eyelashes and mouth; and 2) skin, nose, and eyebrows into different consistency combinations. Their findings indicated that increased inconsistency in human realism will evoke discomfort. They also found that stimuli that are the most difficult to categorize were not rated the eeriest and thus claimed that categorization uncertainty might not explain the uncanny valley concept.

3.1.3 Emotional Effects of The Uncanny Valley

It is never possible to create a virtual agent without considering facial expressions and the effect of emotions. The experience of uncanny valley varies from individual to individual, and emotions play a critical role in attenuating or amplifying such experience.

Tinwell, Grimshaw, Nabi and Williams (2011) investigated whether a lack of facial expressions in the upper part of the face would increase the “uncanniness” of animated
characters. Their findings revealed that angry and fearful expressions on an animated character were rated comparatively lower in familiarity than human characters, whereas sadness favored the transition from human to animated agents in terms of familiarity and human likeness. They explained that this is due to when sadness is portrayed by a virtual character; people tend to project empathy or sympathy to the character, which alleviate the uncanny feeling of viewers. Tinwell et al. (2011) concluded that viewers’ uncanny experience varies with different expressions worn by a virtual agent. A later study by Tinwell and Sloan (2014) recruiting children between 9 to 11 years old reported similar findings which avatars portraying happiness expressions were rated less eerie and friendlier than those with startled expressions. Contrary to happiness, which might ease children’s feeling as a positive emotion, a startled expression created apprehension and confusion (Tinwell et al., 2011; Tinwell & Sloan, 2014). This could possibly explain why happiness was less uncanny than startled expressions. On the contrary, Mathur and Reichiling (Mathur & Reichiling, 2016) had participants viewed and rated 80 robot faces varying different human likeness levels and found that emotions did not influence human likeness level, but positive emotion did associate with higher likeability score. The effect sustained even with low-emotion faces.

The conflicting results could derive from several possible reasons. First, facial expressions are often considered to be based on seven basic emotions (Ekman, 2003), consisting of both positive, negative, and neutral (e.g., surprise) expressions. Some emotions are easier to recognize; for example, happiness was recognized with high accuracy even with low intensities (Hess et al., 1997), while people have a tendency to confuse between anger and disgust, or surprise and fear (Gagnon, Gosselin, Hudon-ven der Buhs, Larocque, & Milliard, 2010). Therefore, it is plausible to assume that an avatar portraying some emotions may fall into or exacerbate the uncanny valley more easily compared to the same avatar wearing other poorly recognized emotions (e.g., anger and disgust). Second, the intensity of an expression
could also have an impact on viewers’ perceptual experience of an avatar. Expressions with higher intensities might trigger more eeriness than those with lower intensities, or vice versa. Alternatively, it might be non-linear, such that facial expressions trigger eeriness at a certain level of intensity only. We lack a sufficient number of studies examining the effects of facial expression of emotion with different intensities on the uncanny valley to draw a firm conclusion from previous studies whose results were inconclusive. Third, virtual agents employed in different studies vary extensively in terms of human likeness and affinity. Given that most studies generated their avatars using software like FACSGen (Roesch et al., 2011) or Autodesk Maya 3D (Autodesk, 2005), the appearance of the avatars is very “animacy or cartoonish” and does not pay meticulous attention to details like skin, or hair. Thus, when observers viewed the relatively lower authentic avatars with emotions superimposed, we are not able to tell if the strangeness is derived from the avatars or their emotions.

3.1.4 Psychophysiological Studies of the Uncanny Valley

Studies using cognitive neuroscience techniques like EEG or ERP to investigate the uncanny valley are still largely absent in the literature.

As discussed in Chapter 2 (session 2.1.3.2), the face-selective N170 is sensitive to different faces, reflecting functional experience in the neural response. It is typically modulated by less commonly seen faces, such as in the “other race” effect. Wiese, Kaufmann and Schweinberger (2014) recruited both Caucasian and Asian participants and tested recognition memory in both groups. They found a more negative N170 for faces from other-race faces in both groups, indicating the structural encoding of faces at an early stage. The well-known inverted face effect also exemplifies atypical faces. Accordingly, larger N170 amplitudes are also observed for inverted faces compared to upright (Sadeh & Yovel, 2010). Taken together, these results indicate that the N170 is modulated by deviation in faces from what people commonly experienced in life. In other words, the N170 is not only sensitive to “shallow”
processing of faces, as in discriminating faces vs. objects; it is also implicated in higher-order perceptual and cognitive mechanism of face processing (Olivares, Iglesias, Saavedra, Truiillo-Barreto, & Valdes-Sosa, 2015).

Wheatley, Weinberg, Looser, Moran and Hajcak (2011) examined the vertex positive potential (VPP; which is believed to be the same component as the N170, Joyce & Rossion, 2005) in response to human faces, doll faces and non-face objects (clocks) and found both human faces and doll faces elicited a larger positivity of the VPP than clocks while no amplitude differences were observed between human and doll faces. They argued that the N170/VPP is sensitive to faces but does not distinguish between human and doll faces. Their findings also revealed a sustained later positivity 400ms post-stimulus for human faces only. Thus, they suggested at least two processing stages were involved in face perception. Faces are shallowly and rapidly processed in the first stage, and the later stage may suggest a perception of the mind. They further argued that these two processes happen automatically but in different time windows. This is supported by a later study, which Looser, Guntupalli and Wheatley (2013) found the detection of global facial form would activate the inferior occipital gyrus while the lateral fusiform gyri and right superior temporal sulcus are sensitive to detect animacy of a face. On the other hand, Balas and Koldewyn (2013) reported that animacy did not modulate the N170 either in latency or amplitude. It indicated that faces would be prioritized to configurable process and configurable and later determined if they are “alive” or possess “the mind” (Gray & Wegner, 2012) and need further cognitive resources.

Mühlberger et al. (2009) compared the P100 and the N170 between artificial emotional faces and human faces, and reported larger P100 and N170 amplitudes for artificial faces. Similarly, Schindler, Zell, Botsch and Kissler (2017) also specified that the process of highly cartoonized or they called stylized faces involved more structural analysis that activates the occipital face area, while the fusiform face area responds to realistic faces. In Balas and
Koldewyn’s (2013) research, they also reported a larger P100 for animated faces than human faces. This could suggest that in early stage of processing, our brain can differentiate between artificial and human faces (Wang, Lilienfeld, & Rochat, 2015) An fMRI study by Shultz and McCarthy (2014) showed that the fusiform gyrus (FG) and posterior superior temporal sulcus (pSTS) were sensitive to the animacy of a virtual agent. More specifically, the FG responded strongly to animate characters than inanimate ones. Although the FG is generally recognized as the structural source (or at least one of the structural sources) of the N170 (Deffke et al., 2007), stimuli in the study were full-body characters that involve very limited detail of a face; therefore, we can only infer that animacy might modulate the N170.

The occipitotemporal P200 has been reported to be modulated by face typicality (Schweinberger & Neumann, 2016). That is, less typical faces tend to delay and attenuate the P200 compared to typical faces (Latinus & Taylor, 2006). Another study also found that P200 amplitude was smaller when faces are more deviated from an “average” face (Zheng, Mondloch, & Segalowitz, 2012). When synthetic faces are involved in comparison to photographic faces, they become “atypical” and may modulate the P200.

Urgen et al. (2015) focused on a later component N400 which is known to respond to violation of predictions. Their stimuli were three video clips consisted of a robot with biological appearance and motion, an android with biological appearance and mechanical motion, and a human with biological appearance and motion. Their results showed that N400 amplitude is larger for the android that has incongruence between motion and appearance. They suggested the violation of predictions could cause the uncanny valley experience.

The abovementioned studies, although investigated some ERP components for artificial faces, it appears that they did not directly measure the experience of the uncanny valley. This
is probably due to the fact that the artificial faces used in these studies are not realistic enough to induce the uncanny valley effect.

3.1.5 Efforts to Cross the Uncanny Valley

While researchers devote much effort to understand the origins of the uncanny valley, another group of people including computer engineers, animators, robot designers, etc. work hard on crossing the valley, and creating hyper-realistic avatars that do not make viewers feel uncomfortable. Recently, NVIDIA created photorealistic level of Hollywood celebrities’ images (see Figure 3.5) using what they referred as Generative Adversarial Network (GAN) with GPU-accelerated machine learning technologies (Karras, Aila, Laine, & Lehtinen, 2018, April-May). This set of images were complemented by media as passing the uncanny valley test. For instance, Forbes magazine commented on the program as a promising attempt to bridge the uncanny valley since machines start learning to make more life-like characters (Altavilla, 2017). The machine-rendered faces reached a new level of photorealism, but the small image size is an issue with NVIDIA’s technology. Springfield Daily (2018) reported the images as “the first seed for a new era in the robotics and artificial intelligence industry”. With these developments in the machine learning techniques, researchers can foresee a very promising step to bridge the valley.
Another remarkable event that should not pass unnoticed is the announcement of the world’s first robot citizen (Stone, 2017). Sophia, a female robot, has become a citizen of Saudi Arabia on October 25, 2017. Her designer, David Hanson, designed her to look like Audrey Hepburn and Sophia can express feelings. Unquestionably, the “citizen Sophia” opens up a whole new set of questions like robotic rights, citizenship, and how she integrates into human’s life. On the other hand, her citizenship could signal the start of a new era, in which human beings begin to accept robots as one of us.

However, we also need to be aware that the work on traversing the valley might not end up in the place we desire. Tinwell et al. (2011) boldly proposed that we might never be possible to cross the valley. Despite pursuing realism on synthetic avatars thanks to the advance of computer graphics, viewers’ ability to discriminating subtle changes from human develops concurrently. Movies, video games, and robotic engineers produce amazing quality of virtual characters every now and then. People, on the one hand, were fascinated by how human those
characters can achieve, yet, on the other hand, rapidly detected even the tiniest difference or breakdown of the avatars. Indeed, for an avatar to be accepted as human, it perhaps has to match not only affinity and human likeness as Mori suggested, but also facial expressions, gaze, voice, gestures, and movements. To integrate all the pieces and put the pieces in the right place like finishing a puzzle seemingly becomes an endless and challenging work, not to mention to escape the almost flawless inspection of our visual system. Just like Tinwell et al. (2011) put forward, we might face the “uncanny wall”.

3.1.6 The Current Study

Earlier work to understand the uncanny valley focused predominantly on behavioral studies. The researchers used self-report questionnaires that seem to be insufficient since they assume the uncanny valley was a conscious and reportable phenomenon, and they omit the possibility of a contribution of unconscious cognitive, perceptual, or emotional processes. Scale rating studies also (over-) simplify the phenomenon into categorical concepts like eeriness, likeness, or pleasantness and respondents made forced choices among the words given. Moreover, Urgen et al. (2015) pointed out that these studies usually involved asking for explicit responses like rating humanness that might be too simplistic to understand people’s experience of the uncanny valley. Further, behavioral methods only measure participants’ experience after viewing stimuli. In the current study, I used EEG that allows to record brain activities during the processing of stimuli, which also has the advantage in avoiding response bias like a questionnaire.

In many studies reviewed above, morphed stimuli are widely used in researching the uncanny valley. Although morphed stimuli between human faces and avatar faces contribute to more controllable experimental stimuli, researchers should be aware of the potential consequences of using those stimuli. First, computer avatar faces today are not simply a morph between human faces and avatar faces. Instead, they are made through computerized simulations of
the facial structure and musculature based on sophisticated algorithms. Therefore, the stimuli might not be representative enough to evoke an uncanny-valley response. Second, morphed faces might be skewed or distorted compared to real computer-generated avatar faces, which could result in feelings of uncanniness or unease (Katsyri et al., 2015). Third, according to Mori (1970) who illustrated robots falling into the valley exhibit higher level of human likeness, in many studies, the morphed stimuli were insufficient to attain the amount of realism that pushes the stimuli to the valley.

In my study, I introduced the Auckland Face Simulator (the AFS) and used it to generate the stimuli. In short, the advantages of using images generated by the AFS as stimuli are: a) it reflects facial muscle movements based on FACS and produces faces and facial expressions closer to their production mechanism; b) the contractions of individual facial muscles or groups of muscles can be controlled by a user or experimenter, providing a set of controllable stimuli ranging from static to dynamic; and c) the AFS contains realistic simulations of the facial musculature, postural muscles of the neck, skull, soft tissue, and skin (Sagar, et al., 2014) and achieves a sufficiently high level of realism to study the uncanny valley.

To summarize, although after Mori popularized the uncanny valley theory in the 1970s, it was not widely researched until the most recent decade or two. Due to the incredibly rapid development of artificial intelligence that pervades our lives, researchers want to know if we are ready to embrace artificial agents. The research of the uncanny valley emerges again and surges rapidly in recent years.

Various explanations and inconsistencies of findings associated with the uncanny valley were rooted in the lack of a firm and precise definition of the two dimensions originated by Mori. To date, there are no clear definitions of either human likeness or affinity. In this study and subsequent analysis, I did not focus on theoretical development. I was, on the other hand,
interested in investigating the psychophysiological manifestation of the uncanny valley. I was unable to find previous studies investigating ERP components of interest such as P200 and FcEP as measures of the uncanny valley. Therefore, I would like to fill this gap in the literature.

To date, we know very little about how our brains process computer avatars, whether realistic or not. Given that computer avatars could start replacing traditional stimuli in psychological studies due to better stimulus control, it is utterly essential to pinpoint what aspects of an avatar face trigger people’s rejection or eeriness towards such stimulus. The current research begins to address the underlying electrophysiological mechanism of computer avatar faces processing and explore the effect of emotional expressions and emotional intensity levels of the perceived realism in avatar faces.

I posed the research questions and hypotheses as below:

RQ3.1) Do our brain respond differently to human and avatar faces? In other words, are there any psychophysiological signatures that support or invalidate the uncanny valley?

\[ H3.1 \]: Which specific ERP components will differ between human and avatar faces in terms of amplitude.

RQ3.2) Do expressions of emotion, or variation in their intensities interact with the uncanny valley effect?

\[ H3.2 \]: Happiness will reduce the eeriness in avatars, and anger will have the opposite effect. Facial expressions with higher intensities will overall increase the eeriness.

For the current experiment, I applied a modified version of the paradigm for the EEG recording used in the previous experiment in Chapter 2. This will be detailed in the next session (3.2). A survey questionnaire was included in the current study to measure perceived
human likeness of the stimuli and correct responses of facial expression recognition. I developed the questionnaire instead of using the Godspeed Questionnaire (Bartneck et al., 2009) for four reasons. First, the Godspeed Questionnaire consists of semantic differential items. In this questionnaire, I asked more straightforward questions like “does the emotion depicted by the face look ‘real’ to you?” Second, participants were forced to choose between two items in the Godspeed Questionnaire, e.g., “ordinary – supernatural”. The pitfall is that respondents cannot respond if their options are not within the dichotomous scale (Krosnick & Presser, 2010). Therefore, I used a 10-point scale to represent a continuum progressing from one end to the other. As such, respondents can rate themselves more accurately. Third, the Godspeed Questionnaire did not cover the dimension of “familiarity” or “affinity” of the uncanny valley. Although in the current questionnaire, I did not have a question specifying familiarity; instead, I asked a question relating to empathy. The level of empathy an observer towards a robot or an avatar face is associated with the observer’s feeling of the human agent (Gonsior et al., 2011). Therefore, I included the question to reflect the dimension of familiarity. Moreover, there are also questions covering human likeness. Fourth, I developed a rather simple and short questionnaire because my experiment is not heavily behavioral in nature, it is rather psychophysiological. I used the questionnaire to understand if there is any connection with the ERP results and to divide groups for further study (see Chapter 5).
3.2 Method

3.2.1 Participants

54 participants (34 females) were recruited through a mailing list and the SONA system, mean age is 21.37 years ($SD = 5.08$). Participants had normal or corrected-to-normal vision and no history of epilepsy or migraine. Moreover, those who have difficulties in recognizing faces or emotional expressions were also excluded. They were compensated with either $20 vouchers or course credits. The study was approved by the University of Auckland Human Participants Ethics Committee.

3.2.2 Materials

**Stimuli.** I used both the Montreal Set of Facial Displays of Emotion (Beaupre & Hess, 2005) and faces generated by the AFS as stimuli. Compared to Chapter 2, I replaced two male photo pictures in the previous stimulus set with two simulated faces (both female models) from the AFS. Therefore, I had two female photo faces and two female simulated faces as stimuli. As the photo faces set was in black and white, I also adjusted the simulation models into black and white. To match with the static photo stimuli, I screen-shot the faces of the AFS models after the required facial expressions were adjusted.

**Survey.** There are six questions in total, asking participants to rate on a scale from 1 (Not at all) to 10 (Very much) regarding the questions. Question 5 consists of two parts. The first part was the same as other questions with a rating scale. In the second part, participants were told to identify the expression on the stimuli. The survey is included in Appendix A.

3.2.3 Procedure

The experimental procedure (see Figure 3.6, below) is largely similar to Chapter 2. Some modifications include: 1) each participant was required to complete 900 trials (approximately 1 hour of EEG recording); 2) participants were required to respond to a surprise expression,
which was different from Chapter 2 (participants would respond to a disgust expression) because the AFS produces more authentic surprise faces than disgust ones; and 3) participants were required to complete a questionnaire on the computer after the EEG session. The EEG recording parameters and data pre-processing were the same as Chapter 2.

In the questionnaire, I used the same stimuli (30 pictures of faces in total) as the EEG task and randomized them. Each face was presented on the screen once for 300ms, preceded by a black cross as fixation for 500ms and followed by six questions after each presentation. Participants saw the image once and were required to rate on a scale of 1 to 10 with “1” (Not at all, representing least agree/humanness) and “10” (Very much, representing most agree/humanness) according to what the question is asking by pressing the number key on the keyboard accordingly (“0” key represents number “10”). The process was self-paced. The next question will appear on the screen once participants have responded to the previous one. The entire questionnaire took about 15 minutes to complete.
Figure 3.6 Experimental procedure of the study and the stimuli of the experiment
3.3 Result

3.3.1 ERP Results

The analysis strategy was similar to that used in Chapter 2. As in the previous experiment, I analyzed the P1, N170, and FcEP (early and late I&II) components of the ERP. To investigate any differences between the processing of simulated face images and photographs of faces, I ran a 2 x 2 x 2 x 3 repeated-measures ANOVA, with electrode (P8 and P8 clusters for the P1 and the N170), realism (photo and simulation), emotion (happiness and anger) and intensity (0% (neutral), 20% and 40%) as factors. I also included the posterior P200 (170-280ms), into the analysis, as the component has been reported to be sensitive to the “typicality” of a face (Schweinberger & Markus, 2016), and seemed a viable candidate to be sensitive to “uncanny valley” responses to simulated faces. I measured the P200 at O1, Oz, and O2, where it was at maximum. The repeated-measures ANOVA for the P200 is 3 x 2 x 2 x 3, which has three electrodes (O1, Oz, and O2) for the first factor. The rest factors were the same as other components.

As in the previous experiment, the amplitudes of each component were measured by taking the average in specific time windows for the ERP components of interest. The details of the components analyzed were shown in Table 3.1. ERP waveforms for the simulated and photographed face images from representative scalp locations are shown in Figure 3.7 (below).
Table 3.1

*Details of Components Analyzed*

<table>
<thead>
<tr>
<th>Component</th>
<th>Time window</th>
<th>Electrode clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>85-110ms</td>
<td>P7 and P8 clusters at posterior sites</td>
</tr>
<tr>
<td>N170</td>
<td>120-160ms</td>
<td>P7 and P8 clusters at posterior sites</td>
</tr>
<tr>
<td>P200</td>
<td>170-280ms</td>
<td>O1, Oz and O2 at occipital sites</td>
</tr>
<tr>
<td>FcEP Early</td>
<td>120-160ms</td>
<td>Fz cluster at frontal site</td>
</tr>
<tr>
<td>FcEP Late I</td>
<td>210-295ms</td>
<td>Fz cluster at frontal site</td>
</tr>
<tr>
<td>FcEP Late II</td>
<td>295-380ms</td>
<td>Fz cluster at frontal site</td>
</tr>
</tbody>
</table>
**Figure 3.7** Grand average waves of photo and simulation conditions. Shaded areas represent ERP components of interest (P7 and P8 clusters: the P1 and the N170; POz: the POz Positivity; O1, Oz and O2: the P200).
**P1 component.** No main effects were observed. There were two significant interaction effects. Electrode and realism returned significant, $F(1, 53) = 6.504, p = 0.014, \eta^2_p = 0.109$, which pairwise comparison with Bonferroni showed that the P1 component reacts to simulated faces larger than photo faces (0.826 $\mu$V vs. 0.700 $\mu$V, $p = 0.012$) at the right-hemisphere electrodes. The interaction between electrode and intensity also reached significance, $F(1, 53) = 5.467, p = 0.007, \eta^2_p = 0.093$. All intensities elicited larger P1 amplitude at the right hemisphere electrodes, but the follow-up comparisons revealed these differences were non-significant, all $ps > 0.05$. Scalp distributions, waveforms, and average voltages in each condition for these two interactions effects are presented in Figure 3.8.
Figure 3.8 a) Individual scalp distributions for P1 component (85-110ms) of all conditions: from top to bottom are 0% (neutral), 20% and 40% facial expressions for both photo and simulation conditions. Topographic maps on the left depict photo conditions and on the right show simulation conditions respectively. b) Waveforms from left and right portray grand averages of photo and simulation conditions at the P7 (left) and P8 (right) electrode clusters from posterior sites. The shaded area on the waveforms reflects the P1 component. c) Bars display the amplitude of the P1 for anger and happiness with each intensity for photo and simulated conditions at left (P7 cluster) and right (P8 cluster) scalp locations. Error bars show SEM.

**N170 component.** There was a main effect for electrode, $F(1, 53) = 4.547, p = 0.038, \eta^2_p = 0.079$. The right hemisphere generated larger N170 in comparison to the left hemisphere (-0.25µV vs. -0.02µV). Realism also returned a significant main effect, $F(1, 53) = 22.396, p < 0.001, \eta^2_p = 0.297$, in which photo faces triggered larger N170 than simulated faces (-0.223µV vs. -0.047µV). The two factors revealed a two-way interaction, $F(1, 53) = 4.979, p = 0.03, \eta^2_p = 0.086$. A test of simple effect with Bonferroni adjustment revealed that N170 amplitude was larger at the right hemisphere for photos (-0.079 µV vs -0.367 µV, $p = 0.014$) while the lateralization effect for simulated faces did not reach significance (0.040 µV vs. -0.133 µV, $p = 0.117$). Figure 3.9 displays these effects.
Figure 3.9 a) Individual scalp distributions for N170 component (120-160ms) of all conditions: from top to bottom are 0% (neutral), 20% and 40% facial expressions for both photo and simulation conditions. Topographic maps on the left depict photo conditions and on the right show simulation conditions respectively. b) Waveforms from left and right portray grand averages of photo and simulation conditions at the P7 (left) and P8 (right) electrode clusters from posterior sites. The shaded area on the waveforms reflects the N170 component. c) Bars display the amplitude of the N170 for anger and happiness with each intensity for photo and simulated conditions at left (P7 cluster) and right (P8 cluster) scalp locations. The curved brackets indicate significant main effects of electrode and realism. *p < 0.05. Error bars show SEM.
**P200 Component.** I observed a laterализation effect of P200, $F(2, 106) = 6.898, p = 0.002, \eta^2_p = 0.115$, which bigger positivity was found on the right side of the occipital area.

Inspection of pair-wise comparisons showed P200 at O2 (2.615µV) is significantly larger than O1 (2.257µV) and Oz (2.271µV) ($ps < 0.05$) and there is no systematic difference between O1 and Oz ($p > 0.05$). I also found a strong main effect of realism $F(1, 53) = 57.155, p < 0.001, \eta^2_p = 0.519$, in which photo faces elicited a much larger P200 than simulated faces (2.619µV vs. 2.144µV).

There was are a two-way interaction between emotion and intensity, $F(2, 106) = 3.307, p = 0.04, \eta^2_p = 0.059$ and a three-way interaction between electrode, emotion and intensity, $F(4, 212) = 2.827, p = 0.026, \eta^2_p = 0.051$. For a more interpretable result, I broke down the three-way interaction into three two-way interactions based on electrode. At O1, I observed a marginally significant main effect of emotions, $F(1, 53) = 3.98, p = 0.051, \eta^2_p = 0.07$, and a significant interaction between emotion and intensity, $F(2, 106) = 4.654, p = 0.012, \eta^2_p = 0.081$. It seems that 40% expression was driving the effect where 40% anger is significantly larger than 40% happiness (2.336µV vs. 2.146µV). At Oz, a main effect of emotions emerged, $F(1, 53) = 5.681, p = 0.021, \eta^2_p = 0.097$, where anger showed more positivity than happiness (2.309µV vs. 2.234µV). There is a significant interaction between emotion and intensity. Again, 40% anger is significantly larger than 40% happiness (2.340µV vs. 2.154µV). At O2, there is a main effect of emotions which showed anger has a more positive amplitude for the P200 than happiness (2.657µV vs. 2.574µV), $F(1, 53) = 5.501, p = 0.023, \eta^2_p = 0.094$. The scalp distribution, waveforms and bar graph of these effects are shown in Figure 3.10.
Figure 3.10 a) Individual scalp distributions for P200 component (170-280ms) of all conditions: from top to bottom are 0% (neutral), 20% and 40% facial expressions for both photo and simulation conditions. Topographic maps on the left depict photo conditions and on the right show simulation conditions respectively. b) Waveforms from left and right portray grand averages of photo and simulation conditions at the O1 (left), Oz (middle) and O2 (right) electrode from occipital sites. The shaded area on the waveforms reflects the P200 component. c) Bars display the amplitude of the P200 for anger and happiness with each intensity for photo and simulated conditions at the O1 (left), Oz (middle) and O2 (right) electrode scalp locations. The curved brackets indicate significant main effects of electrode, realism and emotion. *p < 0.05. Error bars show SEM.
**FcEP late I.** I found significant main effects on realism, $F(1, 53) = 49.353, p < 0.001, \eta^2_p = 0.482,$ and emotion, $F(1, 53) = 5.486, p = 0.023, \eta^2_p = 0.094.$ Simulated faces showed larger positivity than photo faces (-1.879µV vs. -2.17µV). Happiness also showed more positivity on the component (-1.993µV vs. -2.057µV). A marginally significant interaction effect between realism and intensities was also found, $F(1, 53) = 3.097, p = 0.058, \eta^2_p = 0.027.$ Pairwise comparison with Bonferroni adjustment showed that simulated faces evoked a larger FcEP across all intensity levels ($ps < 0.05$) than photo faces, whereas no systematic differences within intensities reached the significant level ($ps > 0.05$). This indicates that the realism differences were driving the interaction effect. Figure 3.11 displays the effects.
Figure 3.11 a) Individual scalp distributions and waveforms for the FcEP late I (210-295ms) of all conditions: from top to bottom are 0% (neutral), 20% anger, 40% anger, 20% happiness, and 40% happiness for both photo and simulation conditions. Topographic maps on the left depict photo conditions and on the right show simulation conditions respectively. b) The waveform portrays grand averages of photo and simulation conditions at the Fz electrode clusters from anterior sites. The shaded area on the waveform reflects the FcEP late I. c) Bars display the amplitude of the FcEP late I for anger and happiness with each intensity for photo and simulated conditions. The curved brackets indicate significant main effects of emotions and realism. *p < 0.05. Error bars show SEM.

FcEP late II. No main effects on factors reached significance. Yet I found a two-way interaction between realism and emotions, $F(1, 53) = 4.275, p = 0.044$, $\eta^2_p = 0.075$, and a three-way interaction between realism, emotion and intensity, $F(1, 53) = 3.471, p = 0.04$, $\eta^2_p = 0.061$. I then performed the analysis of simple effects with Bonferroni correction and found that realism effect was significant for 40% anger, namely FcEP responded significantly larger for anger 40% of photo faces than simulated faces (-1.387 $\mu$V vs. -1.583 $\mu$V, $p = 0.026$).
Furthermore, at 40% intensity level, happiness also elicited a larger FcEP than anger (-1.371 μV vs. -1.583 μV, p = 0.030) for simulated faces.

I also broke down the three-way interaction into two simple two-way interactions based on emotions. In the anger condition, 40% anger elicited significantly more positivity in the photo faces compared to simulated faces, whereas both 0% and 20% expressions were more positive in simulated faces than photo faces, $F(1, 53) = 3.881, p = 0.035, \eta^2_p = 0.068$. In the happiness condition, I observed a main effect in realism, namely simulated faces yielded more positivity than photo faces, $F(1, 53) = 7.264, p = 0.009, \eta^2_p = 0.121$. These effects are shown in Figure 3.12.
Figure 3.12 a) Individual scalp distributions and waveforms for the FcEP late II (295-380ms) of all conditions: from top to bottom are 0% (neutral), 20% anger, 40% anger, 20% happiness, and 40% happiness for both photo and simulation conditions. Topographic maps on the left depict photo conditions and on the right show simulation conditions respectively. b) The waveform portrays grand averages of photo and simulation conditions at the Fz electrode clusters from anterior sites. The shaded area on the waveform reflects the FcEP late II. c) Bars display the amplitude of the FcEP late II for anger and happiness with each intensity for photo and simulated conditions. Error bars show SEM.
3.3.2 Self-report Questionnaire Result

3.3.2.1 Human Likeness

I calculated the human likeness scores of photo and simulated faces by averaging scores of five questions from 53 participants (one was removed as an outlier). I ran a repeated-measure ANOVA to investigate how models (photo and simulation), emotions (anger and happiness) and intensities [neutral (0%), 20% and 40%] affect the human likeness score. Because the neutral stimuli lack any emotion, I arbitrarily assigned each to one of the two emotion categories (i.e., anger-neutral and happiness-neutral) as factors for ANOVA. The human likeness scores of photo and simulated faces showed no systematic difference ($p > 0.05$).

Participants rated the simulated faces as real and human as photo faces. Correlational test also supported the idea ($r(53) = 0.635, p < 0.001$), as shown in Figure 3.13. There is a tendency that those who rated photo faces are more real and human also gave high scores for simulated faces and vice versa. Emotions also showed no difference. There was a significant main effect of intensities, $F(2, 104) = 10.902, p < 0.001, \eta^2_p = 0.173$. 40% expressions obtained the highest score in human likeness compared to 20% expressions and neutral, confirmed by post-hoc comparison ($ps < 0.05$); and there was no enough evidence suggesting differences in human likeness score between 20% expressions and neutral.

There was a significant interaction between model and emotion, $F(1, 52) = 10.918, p = 0.002, \eta^2_p = 0.174$. Happiness on simulated faces received a higher score on human likeness than photo faces while there was no systematic difference in human likeness score of anger expression between photo and simulated faces. A marginal significant interaction between model and intensity was also found, $F(2, 104) = 3.15, p = 0.051, \eta^2_p = 0.057$. There was also a significant three-way interaction between model, emotion and intensity, $F(2, 104) = 3.31, p = 0.049, \eta^2_p = 0.06$. To further investigate this interaction, I broke down the three-way interaction into two two-way interactions based on models. The first interaction between
emotion and intensity of the photo model revealed neither main effects on emotion and intensities nor a significant interaction between these two factors. The second interaction between emotion and intensity on simulated faces yielded main effects of emotion ($F(1, 52) = 7.627, p = 0.008, \eta^2_p = 0.128$), and intensity ($F(2, 104) = 10.082, p < 0.001, \eta^2_p = 0.162$).

Happiness is perceived as more human than anger, and 40% expressions also received higher score on humanness compared to neutral (0%) and 20% expressions ($ps < 0.05$), whereas no enough evidence supported a difference between neutral and 20% expressions ($p > 0.05$).

Hence, happiness and higher intensity of expressions tend to contribute to the perceived humanness of computer avatars.
Figure 3.13 a) Scatter plot of photo and simulation indicates human likeness score. b) Bars display the human likeness score of anger and happiness with 3 levels of intensity (0%, 20% and 40%) for photo and simulation models. The curved brackets indicate a significant main effect of intensity. *p < 0.05. Error bars show SEM.
3.3.2.2 Correct Response

I also measured the correct response of recognition of emotions for all stimuli (30 pictures, including 20 photo faces and 10 simulated faces). Generally, participants have a relatively low accuracy rate of facial expression recognition. The overall score was 12.53 out of 30 points, which gave the overall accuracy rate of 41.76%. Again, I ran a repeated-measure ANOVA to investigate whether and how model (photo/simulated faces), emotion (anger and happiness) and intensity [neutral (0%), 20% and 40%] affected the correct responses.

Model, emotion and intensity showed main effects. The correct response rate of photo faces is significantly higher than that of simulated faces, $F(1, 52) = 42.352, p < 0.001, \eta^2_p = 0.449$. Participants obtained 47.45% accuracy rate for photo faces whereas for simulated faces, participants reached 30.38% accuracy rate. Happy faces obtained significantly higher correct response rate than angry faces, $F(1, 52) = 34.87, p < 0.001, \eta^2_p = 0.401$. Happiness received 43.9% correct response rate whereas anger has 32.0% correct response rate. 40% expressions have the highest accuracy rate of 55.0%, followed by neutral expression with 33.3% correct response rate and 20% expressions with 25.7% correct response rate, $F(2, 104) = 27.868, p < 0.001, \eta^2_p = 0.349$. Post-Hoc test showed that 40% expressions are more accurately recognized ($p < 0.05$). The accuracy rate of 0% and 20% did not distinguish from each other ($p > 0.05$).

I also found a significant interaction between model and emotion, $F(1, 52) = 24.562, p < 0.001, \eta^2_p = 0.321$. This indicated that the higher correct response rate for photo faces compared to simulated faces was modulated by emotion. Specifically, anger (20.1%) has extensively lower accurate rate than happiness (40.6%) in simulated faces. Emotion and intensity also returns a significant interaction, $F(2, 104) = 57.02, p < 0.001, \eta^2_p = 0.523$. Both anger and happiness received less correct responses on neutral and 20% expressions, while 40% happiness outperformed and drove the interaction effect. I also found a three-way
interaction between model, emotion and intensity, \( F(2, 104) = 7.812, p = 0.002, \eta^2_p = 0.131 \).

Again, I broke it down into two interactions based on models. For photo faces, I found a main effect of intensities where people recognize 40% expressions better than 20% and neutral ones, \( F(2, 104) = 19.126, p < 0.001, \eta^2_p = 0.269 \), but no main effect of emotions. An interaction between emotion and intensity suggested the correct recognition of 40% happiness was significantly better than all anger expressions and lower intense happiness, \( F(2, 104) = 27.601, p < 0.001, \eta^2_p = 0.347 \). For simulation models, I observed main effects of emotion and intensity. Happiness obtained twice as much as anger in correct responses, \( F(1, 52) = 59.831, p < 0.001, \eta^2_p = 0.535 \). 40% expressions has the highest accuracy rate with 45.8% while 20% has the lowest with 15.1%, and neutral has a 30.2% accuracy rate, \( F(2, 104) = 17.645, p < 0.001, \eta^2_p = 0.253 \). Pair-wise comparisons reflected significant differences between all intensity levels \((p < 0.05)\). A significant interaction between emotion and intensity was reported, \( F(2, 104) = 36.274, p < 0.001, \eta^2_p = 0.411 \). People performed significantly better in recognizing 40% happiness than 40% anger. Furthermore, 20% expressions, especially 20% anger, were generally more difficult to identify than neutral and 40% expressions and happiness expressions. Bar graph of the effects are shown in Figure 3.14.
Figure 3.14 Bars display the correct response rate of anger and happiness with 3 levels of intensity (0%, 20% and 40%) for photo and simulation models. The curved brackets indicate significant main effects of model, emotion and intensity. *p < 0.05. Error bars show SEM.

Taken together, it suggested that higher intensity of emotions contributes to the human likeness and the finding that 40% expressions of simulation are perceived more human than photo faces is intriguing. Further, 40% intensity also contributes to more accurate emotion recognition. 40% happiness is perceived the most human and simultaneously has the highest correct response rate. In contrast, 20% anger has the lowest correct response rate and is perceived as less human.
3.4 Discussion

This exploratory study aimed to explore whether the ERP components of interest (P1, N170, FcEP, and P200) were modulated by the realism of faces, and/or the emotion and intensity displayed. No significant main effects were found for the P1. I found that the realism of faces attenuated N170, especially at left hemisphere electrodes. For the P200, photo faces evoked a larger amplitude than simulated faces. A larger positivity of the P200 was also observed at the right side of the occipital area. Both realism and happiness elicited more positive FcEP late I component. In the FcEP late II component, the perception of realism seemed to depend on different types of emotions.

3.4.1 Survey Findings

Photo and simulated faces did not reach statistical significance in the ratings of human likeness. The current analysis showed that people who perceived photo faces to be more humanlike also responded similarly to simulated faces. I see two possible – and not mutually exclusive – interpretations for this result. First, the simulated faces and real faces are genuinely similarly humanlike in appearance, and people cannot differentiate between the two types of faces. Second, there might be individual differences in the criterion for perceiving a face as humanlike. Some people may have a higher threshold – or a stricter criterion – for accepting a face as “real”. Faces failing to reach the threshold will score low in the human likeness dimension regardless of whether faces are computer-generated or photos, and vice versa.

I also found that 40% expressions increased the human likeness score. Further analysis revealed that both emotion and intensity played a role in the perceived humanness on avatar models, but not photo faces. Human likeness score was in favor of avatar faces with happy expressions compared to anger. In particular, 40% happiness was rated the most humanlike,
while neutral and 20% anger were perceived the least humanlike. This finding is partly consistent with my hypotheses, as I expected that happiness would increase human likeness for avatar faces because positive emotions were reported to associate with higher likeability score (Mathur & Reichiling, 2016). On the other hand, the result also disagrees with Tinwell, Grimshaw and Abdel-Nabi (2011), who found happiness was rated the least familiar and humanlike expression in avatars. It could possibly be that the dynamic stimuli used by Tinwell et al. (2011) contributed to the induction of an “uncanny” feeling when the movement might not be precisely right. To my surprise, avatars displaying both 20% and 40% happiness scored higher than photo faces in human likeness, although for 20% happiness this difference did not reach significance. In other words, simulated faces with higher intensity level of happy expressions were perceived to be more human than photo faces. It may provide some insights for ways to improve robotic design and human-robot interaction. For instance, designers should improve angry expressions to reach a higher level of human likeness or may take advantage of happy expressions as a way to make their creations seem more human. Further, as photo faces in the current study are morphed stimuli, it might indicate that avatar faces possess some advantages as stimuli.

For correct responses of emotional expression recognition, photo faces and happy faces were identified correctly more often than simulated faces and angry faces. The results are not consistent with Joyal et al.’s (2014) and Dyck et al.’s (2008) results. Both studies did not find a significant difference in correct recognition either between virtual and photo faces, or happiness and anger. Explanations are two-fold. First, stimuli in the current study were only presented for 300ms, whereas stimuli were presented for 10s and 7s in the other two studies respectively. Longer presentation time tends to increase the accuracy rate for emotional recognition (Derntl, Seidel, Kainz, & Carbon, 2009). Therefore, the accuracy rates for virtual and photo faces, or happiness and anger were equally high and probably at ceiling. Second, I
used relatively low intensity expressions compared to the aforementioned studies, which employed full intensity expressions. Again, using full intensity expressions eased the difficulty in the recognition task and aids to achieve a high accuracy rate. In my study, I found angry simulated faces, especially 20% anger, received the lowest correct response rate. This result, too, is quite contradictory with the existing literature (Fox et al., 2000; Ruffman, Ng, & Jenkin, 2009), which has found that angry faces are more rapidly and efficiently detected than happiness, suggesting an anger superiority effect (Hansen & Hansen, 1988). However, Becker et al. (2012) argued that stimuli in these studies were simple schematic drawings of anger and that participants may recognize the drawing patterns in the absence of true emotional processing when performing the tasks. They claimed that when stimuli are more realistic such as photographic images, an advantage emerged for happiness in emotion-recognition tasks. The faces used in this study have ecologically valid features; even avatar faces are more realistic than schematic drawings. Hence, the stimuli in the study may have “happiness advantage” in emotional recognition (Becker et al., 2012), and may explain my finding of higher accuracy rate for happiness than anger.

3.4.2 ERP Findings

**Lateralization for the P1 and the N170.** For ERP components, I obtained lateralization of the P1 and N170 at the right hemisphere. Both components were larger in amplitude at the right hemisphere electrodes. P1 was larger to simulated faces whereas N170 was larger for photo faces. This is in line with many other studies, in which right-hemisphere dominance for face processing – and the amplitude of N170 in particular – has been consistently observed (Calvo & Beltran, 2014; Rossion, 2014).

**P1.** Halit et al. (2000) reported that P1 is sensitive to facial configurational change. They presented stretched human faces which the eyes were moved 20%-30% upwards of the distance between pupil and tip of nose, and the mouth was moved downwards of the same
portion of the distance between the tip of nose and tip of the lip. They found enhanced P1 amplitude for stretched faces, which suggested a rapid top-down process of faces or attentional modulation. In the current findings, I have some evidence that P1 amplitude is larger for simulated faces, but in an interaction effect with lateralization. I offer two possible explanations. One is that low-level visual features like contrast, luminance, etc, have been shown to modulate P1 amplitude (Batty & Taylor, 2003). Stronger stimuli (higher contrast, higher luminance) tend to evoke a stronger P1. Furthermore, stimuli that are attended also evoke a larger P1 amplitude than stimuli that are unattended, but typically only when they are in the peripheral visual field (Fu, Fan, Chen, & Zhuo, 2001). In the current study, stimuli were presented at the center of the visual field. Therefore, it is unlikely that either of these explanations hold for the P1 effect here. The other plausible explanation is that I used computer-generated faces which could be more difficult to compare to the prototype (namely the photo faces) in such an early stage. The P1 effect may be postponed or merged with the next component, N170. Valentine (1991) outlined that faces were encoded multi-dimensionally, and each dimension served the purpose of face discrimination. Therefore, P1 might respond to identifying these dimensions and assist in the later discrimination stage between faces.

N170. The N170 is a face-sensitive component, as discussed in many papers and my previous chapter. The N170 typically manifests as an enhancement of the visual N1 component when a stimulus is a face, compared to other, non-face objects or stimuli (Batty & Taylor, 2003). Compared to photo faces, simulated faces are likely to be perceived as less “facey”. Our brain can always detect some details that do not match a real human face precisely. Simulated faces are not perfect enough to escape such a detailed inspection. Evidences are suggesting that a new face is processed and encoded by comparing to a prototype that represents an average face (Leopold, Toole, Vetter, & Blanz, 2001; Levin, 2000). We will then respond to the
deviation of features from the prototype. As a result, our brain may conclude that simulated faces are not “facey” enough and respond to an attenuated N170. Apart from that, studies have suggested that high working memory load tended to attenuate N170 amplitude (MacNamara, Schmidt, Zelinsky, & Hajcak, 2012). Simulated faces with a smaller N170 may indicate that something unrealistic about the avatar faces consume a higher working memory load.

I found no significant effects of emotion and intensity for the N170, which is inconsistent with the findings in Experiment 1 (see Chapter 2) in which the N170 responded greater to angry expressions than happy ones. It is interesting to speculate that perhaps when we are processing a highly realistic face, we draw a majority of our attentional and cognitive resources to deal with the face. Compared to Experiment 1, the emotional effect appeared in the N170 time window, whereas in the current analysis, the effect was observed in the P200 time window. As a result, I speculate that the process of emotions or intensities may either be postponed or inhibited when realism of a face is getting in the way.

**P200.** Another face-sensitive component subsequent to the N170 emerged at the occipital temporal area is the P200. I found a larger P200 amplitude for photo faces than simulated faces. Moreover, the P200 is larger at right hemisphere electrodes. The current result is consistent with Halit et al. (2000) as they also reported a larger P200 for intact faces than artificially stretched faces. They argued that the P200 is affected by the typicality of a face (Wuttke & Schweinberger, 2019). Therefore, when viewing a computer-generated face, our visual system compares the face with a human face and responds to faces that deviate the least to the human face.

Schindler et al. (2017) asserted that more realistic faces might be perceived as having a unique identity, which consumes more social cognitive resources that involve individual
identity representation. This might as well explain a larger P200 of photo faces in comparison to simulated faces, as people project social identities to photographic faces than computer-generated faces. Further, Halit et al. (2000) suggested that the P200 is sensitive to the variation of typicality only within a certain identity. The current findings disagreed with Halit et al.’s (2000) claim since stimuli in this study consisted of four identities (two for photo faces and two for avatar faces), and there was a strong P200 effect. This might be because compared to Halit et al. (2000), stimuli in this study involved more dimensions of face typicality. Halit et al. (2000) changed the configuration between features like eyes and mouth, the typicality majorly refers to the distance or space between these features. In my study, participants not only inspected the configuration of features between photo and avatar faces but also examined at the realism dimension, which the typicality will then consist of both spatial configuration and realism variations. Consequently, the process of identity becomes secondary, thus the P200 responded to face typicality sustained across different identities.

A line of evidence supported the P200’s sensitivity on facial prototypicality in the other-race effect. Stahl, Wiese and Schweinberger (2008) reported an increase P200 for own-race faces suggesting second-order spatial configuration and differences in processing faces of own race and other races. In the same vein, the larger P200 response in the current study indicated our brain may treat avatar faces as if they are other-races or otherwise atypical. Additionally, Stahl et al. (2008) found that the P200 effect was absent in an “expert” group consisting of participants who engaged intensively with other races. Participants in this study are mostly university students. Given the age, it is reasonable to assume that they are at least familiar with computer-generated human characters or faces. The presence of the P200 in the current study with the expertise level of the participants, I thought that when it involves realism of a face (between a human face and an avatar face), expertise in virtual reality is unable to eliminate the differences between a human face and an avatar face. This could be possibly
due to the process of realism is crucial in differentiating human and avatar faces, and expertise does not lower such importance.

Nevertheless, it should be noticed that the concept of facial prototypicality is not well-defined. By assumption, normal photo faces were used as controlled typicality, however we cannot neglect that human faces vary extensively among each other. Factors like race or skin color can define prototypicality very differently. Hence, it would be interesting to investigate the ERP responses to those atypical human faces without artificial creation, such as scars or even distorted faces due to accidents, compared to normal human faces and avatar faces.

**FcEP.** The FcEP is an emotion-sensitive component recorded at frontal sites. Overall, the FcEP responded to the avatar faces more than photo faces in the late I time window. It suggested that when viewing an avatar face, observers are more “emotional” compared to a photo face regardless of emotions. Eimer and Holmes (2007) emphasized that the frontocentral positivity could signal a rapid detection of emotionally significant stimuli. Mori defined and conceptualized the uncanny valley with emotional content. That is, when we are presented with a synthetic face and realize it is not quite human, according to Mori, we then experience negative emotions (discomfort, eeriness, and etc.). Such responses, may result in the modulation of the FcEP, which could support the hypothesis of emotional involvement in the uncanny valley theory.

Realism, emotion and intensity interacted on the FcEP late II component. Schindler et al. (2017) also observed an interaction between emotion and realism but at a much earlier stage, in the N170 time window. They asserted that such an effect could indicate an interaction between structural decoding and emotional context instead of dual processing of identity and emotion. Moreover, consistent with Schindler et al.’s (2017) findings, the current results indicated for angry expressions, photo faces elicited larger amplitude, while for happiness,
simulated faces have larger ERP impact. It is worth noting that the interaction between emotions and realism in this study emerged at a much later stage than Schindler et al.’s (2017). This is possibly due to a shorter presentation time (300ms), and the use of low-intensity emotions (20% and 40% intensity) in my study which might only allow the brain to pick up the most salient factor of the face (e.g. realism) at an early stage and process less important characteristics (e.g. emotions) at a later phase. Furthermore, participants were passively attending to the faces presented as they were not given tasks and were instructed to attend to the faces in Schindler et al.’s (2017) experiment. Participants in the current study were instructed to respond to surprise faces although their responses to these stimuli were irrelevant to the study. Hence, realism of a face is processed as a priority. Emotions or intensities worn on the face become secondary, and may be postponed in processing. Another possibility for the emotion-realism interaction effect at a later stage could be the use of lower intensities of expressions. The experience of the uncanny valley may overwhelm lower-intensity emotions.

Combining findings from the N170, P200 and FcEP, we can presume that at an early stage, the brain draws upon attentional resources and inspects the avatar face with in the context of a human-face frame from the memory. The avatar face is then compared with an average face, namely the facial prototype. Once the avatar face is identified to be nearly human but imperfect, emotional response kicks in, which signals that the avatar face falls into the uncanny valley. In the current study, the late and not strong emotional effect might raise a second thought about the role of emotion played in experiencing “uncanny”. The very initial concept of “uncanny” by Jentsch (1906) did not emphasize emotional response as Mori did. It is also possible that the emotional signal of being uncanny is fundamentally different to the processing of emotional facial expressions, since in my previous study (see Chapter 2), the
emotional effect presented and was sustained. Of course, it needs more systematic studies in the future to confirm or disprove this assumption.

3.4.3 Comparison between survey and ERP findings

The current ERP results seem to map onto the survey results partially but not completely. For human likeness, the ERP results and survey results produced inconsistent outcomes. Behaviorally, most people did not differentiate avatar and photo faces with regard to realism though individuals recognized facial expressions better with photo faces than simulated faces. Yet psychophysically, the ERP responses demonstrated a large difference between avatar and photo faces. Further, the survey results demonstrated emotional effects on human likeness and correct responses. Happiness increases human likeness of an avatar and is easier to recognize than anger. However, I did not obtain strong evidence of the interaction between emotions and realism from the ERP results. In addition, the effect of intensities on differentiating stimuli in terms of realism and emotion was mostly absent in both behavioral and psychophysiological findings. A plausible reason might be the emotions were not intense enough to generate an ERP effect given the prioritized process of realism. Besides, it is also possible that what you “see” does not equal to what your brain “sees”. This occurs in visual distortions such as the Muller-Lyer effect, which the length of a line is perceived longer when the arrows attached to the each end of the line pointing inward than the arrows pointing outwards (Zhang, et al., 2013). Zhang et al. (2013) reported significant differences in response times of four different illusion stimuli in the behavioral study yet the difference was absent in early ERP components like P1, N170 and N2. They suggested that the processing of distortions might be associated with high-level cognitive factors. Therefore, emotion and intensity effects may also be controlled by high-level cognitive factors compared to realism and emerge at a later latency. Further, it is also important to consider that the ERPs measured in this study reflect the visual processing and early evaluation of the face – pretty much at the
“input” end of processing. Responses are at the “output” end, and may be influenced by factors that are invisible to ERPs. Likewise, ERPs may be sensitive to some aspects of the stimuli that do not significantly influence the behavioral response.

3.4.4 The uncanny valley processing

To some extent, it is plausible to assume that uncanny valley might hold a function that allows human beings to detect when something is not our kind. Thus, it may be particularly significant that the ERP differentiated the avatar faces at a very early stage, possibly initiated as early as the P1 time window, and the response sustained till 300ms. Furthermore, the later occurrence of ERP responses to emotion and intensity compared to the Experiment 1 (see Chapter 2) may indicate the prioritized process of realism. The current results may shed light on evolutionary aspects of understanding the uncanny valley phenomenon. From an evolutionary aspect, computer-generated avatars are new to human beings. We are probably not hard-wired to process avatars. Thus I assumed that when we see a hyper-realistic avatar face which almost passes the inspection of our visual system, we allocate more cognitive resources to process the face. When avatar faces become too difficult to differentiate from human faces, we might activate our defense mechanism and emotionally reject avatar faces. Another aspect that should be noted is that based on what Mori (1970) suggests, the eeriness will diminish if the level of realism goes past the uncanny valley. Nevertheless, we do not know this for a fact since first of all, the technology is not yet available in creating 100% humanlike characters, and secondly, it is still unknown where the tipping point of falling into or getting rid of the uncanny valley will be.

The present study demonstrated that simulated faces that are perceptually undifferentiable have not reached the realism level to replace human face stimuli. In other words, the findings obtained using simulated faces are not completely transferrable to humans (Schindler, et al.,
The current study also indicated that ERP study is a more reliable way to test the realism of computer-generated faces in comparison to self-report questionnaires.

3.4.5 Limitations

I note several limitations of this study. First, I have only a very limited number of stimuli, and these are restricted to two female avatar models. Using a greater range of avatars, and including both male and female models will be necessary to generalize the findings better. Additionally, studies incorporating human avatars or humanoid often face similar constraints of a very low number of stimuli. For instance, two computer-generated avatars were used in Schindler et al.’s (2017) and Shultz and McCarthy’s (2014) studies. Some studies (Ikeda, et al., 2017; Pour, Taheri, Alemi, & Mehdari, 2018; Urgen, et al., 2015) used only one agent as the stimuli. This is because to build realistic computer-generated avatars or humanoids are laborious and tech-intensive, the number of completed avatars that are available for studies is generally small in the field. With the advancement in AI, more avatars are expected to be available for future studies.

On a related note, another limitation of the current study is only photographic and hyper-realistic avatar faces were used without a comparison of not-so-realistic avatar faces that can act as a baseline for the lower end of human likeness. Another aspect is that I did not employ a more thorough questionnaire in the current study although this study concentrated majorly on the psychophysiological aspect. To develop a questionnaire that measures both human likeness and affinity more thoroughly would be necessary to investigate and interpret how individuals perceive highly realistic computer avatars both perceptually and psychophysiological. Moreover, for future studies, I would recommend using dynamic facial stimuli and more intensive expressions to examine the effect of emotional expressions on the uncanny valley.
Chapter 4 Model Differences

4.1 Introduction

In the behavioral study (Chapter 3), I received informal feedback from the participants commenting that they were unable to differentiate Leah from photos, but they could easily identify Xyza as an avatar. Such comments caught my attention, as it was certainly not the intention to make the two models with different levels of realism. Rather, the developer intended to create and design both models as hyper-realistic simulations of human faces. As a result, I was interested in whether - and how - the individual responses to the two simulated face models are similar or different. Thus, I continued to explore how individuals process different levels of realism when they come across computer-generated characters. To the best of my knowledge, studies that discussed and/or compared the levels of realism between avatar models are still absent in the literature. A majority of studies have concentrated on creating or matching avatars with competitive realism level to humans, which is similar to the approach in Chapter 3.

In reviewing the uncanny valley literature and findings from Chapter 3, it is quite evident that we may encounter the uncanny valley even with hyper-realistic computer-generated avatars (Tinwell, Grimshaw, & Williams, 2011). In the previous chapter, I demonstrated that neural signatures (i.e. EEG recordings) be more sensitive to the realism of an avatar than overt human behavior (i.e. “eyeballing” and rating the stimuli). Considering the observations of the participants that the two avatars differed in their “humanness” (even if they received similar ratings in the survey), I was curious to see if the uncanny valley effect presents between avatar models. Is, for instance, one model “uncannier” or “creepier” than the other? In the current analysis I will focus on investigating whether ERPs can provide a more effective means to identify and differentiate “more real” from “less real” avatars. Likewise, examining
the difference in ERP response to different avatars offers the intriguing possibility of
discovering new components or effects related to the experience of the uncanny valley.

Based on my last experiment’s results (Chapter 3), I will scrutinize the N170 and the P200
further by analyzing the responses to Leah and Xyza separately. The N170 reflects how
“face-like” a stimulus is being perceived (Bentin et al., 2007; Eimer, 2000). Thus, although
the avatars are both unmistakably faces, I surmise that the more “humanlike” appearance of
Leah will evoke a larger N170 amplitude than Xyza. For the P200, which responds to face
typicality (Stahl et al., 2008), I similarly hypothesize that Leah will elicit an enhanced P200
compared to Xyza. Results in Chapter 3 suggested that “humanlike” faces are more typical
than computer-generated faces. As such, Leah appears more typical than somewhat artificial
Xyza. Given that, I would anticipate a larger P200 for Leah than for Xyza, and similar P200
for Leah and photos.

It is important to note that this chapter is very much a post-hoc exploration of a small data set
in the hope of discovering some possible avenues for future research. The limited number of
avatars and lack of systematic control of differences between them means that generalizable
results are difficult or impossible to obtain. Nonetheless, this dataset offers a valuable
opportunity for exploration and should generate future research with larger numbers of better-
controlled stimuli.
4.2 Methods

I reanalyzed the data from Experiment 2 to explore differences in ERPs evoked by the two avatars, Leah and Xyza (see Chapter 3 for details of data collection and EEG signal conditioning). Specifically, I compared ERPs evoked by each avatar with one another and with those evoked by photos within a 800ms epoch beginning 200ms prior to the onset of the faces. No new data were collected for this analysis.

4.3 Result

Grand-average ERPs for each face model (Leah, Xyza and photo) are shown for each of 12 electrode clusters in Figure 4.1. Visual inspection of this figure reveals that evoked activity is largely concentrated at parietal and occipital areas. Paired with a closer visual inspection of the topographic scalp maps of the differences between photo and simulated faces (see Figure 3.8-Figure 3.12 in Chapter 3), I detected prominent differences at parietal and occipital electrode clusters between photo and simulated faces in the time windows of the N170 (120-160ms) and the P200 (170-280ms). Additionally, I noticed a salient occipito-parietal positivity distributed around the POz scalp location in the N170 time window. The scalp distribution distinguishes strongly between the three models, with Leah showing a strong positivity centered on POz, followed by Xyza and photo respectively. Inspection of the corresponding waveforms suggests that this central positivity is superimposed on the negative-going activity that constitutes the N170. This suggests that there are at least two simultaneous (or near-simultaneous) responses to these faces. To further explore this “POz positivity”, I measured the average amplitude for each condition (two avatars and photo) in a time window between 120 and 160ms (corresponding to the N170 time window) at a matrix of 12 scalp locations (nine clusters and three single electrodes; see Table 4.1 and Figure 4.2).
I subjected these average amplitudes to a 3x4x3 repeated-measures ANOVA with side (left, middle, and right), anterior/posterior (front, central, posterior, and occipital) and model (Leah, Xyza and photo) as factors. Follow-up pairwise comparisons were adjusted with Bonferroni corrections.

I found significant main effects of side, $F(2, 106) = 13.844, p < 0.001, \eta^2_p = 0.207$, and model, $F(2, 106) = 15.626, p < 0.001, \eta^2_p = 0.228$. Pair-wise comparisons showed that overall the amplitude for the right side electrode cluster was larger than the left and middle sides clusters respectively (-0.002 µV vs. 0.081 µV vs. 0.258 µV) ($ps < 0.05$) while no significant differences were observed between the left and middle clusters ($p > 0.05$). Further pair-wise comparisons showed the amplitude for photos was significantly greater than for the two avatars (respectively 0.215 µV vs. 0.127 µV vs. -0.006 µV; $ps < 0.05$), and the overall amplitude difference between Leah and Xyza was not significant ($p > 0.05$). There was also a significant interaction between side and model, $(F(4, 212) = 4.406, p = 0.003, \eta^2_p = 0.077)$. Post-hoc comparison revealed that photo faces evoked a larger negativity than Xyza and Leah at both left and right electrode sites ($p < 0.05$), whereas Xyza and Leah did not show significant differences in the amplitude ($ps > 0.05$). At middle electrode sites, three models differed significantly from each other, with Leah (0.387 µV) evoked the largest positivity, followed by Xyza (0.260 µV) and photo faces (0.126 µV) ($ps < 0.05$). Most importantly, a significant interaction between side and anterior/posterior in the head matrix analysis exhibited a maximum positivity at the middle posterior location, $F(6, 318) = 4.406, p < 0.001, \eta^2_p = 0.178$, which confirmed the POz cluster as location of interest for this positivity.

There was also a significant three-way interaction between side, anterior/posterior and model, $F(12, 636) = 4.143, p = 0.001, \eta^2_p = 0.072$. The break-down of the interaction is further described below in Model Differences (session 4.3.1.1). Table 4.2 displays bar graphs of
average amplitude in the 120-160 ms time window for all 12 electrodes locations within the N170/POz positivity time window.

*Figure 4.1* ERP waveforms collapsed across all emotional expressions and intensities compared between Leah, Xyza and Photo. The waveforms are corresponding to the electrodes (clusters) presented in Table 4.1.
Table 4.1

Electrodes (clusters) of interests

<table>
<thead>
<tr>
<th></th>
<th>Left</th>
<th>Middle</th>
<th>Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>Front (clusters)</td>
<td>F5</td>
<td>Fz</td>
<td>F6</td>
</tr>
<tr>
<td>Central (clusters)</td>
<td>C5</td>
<td>Cz</td>
<td>C6</td>
</tr>
<tr>
<td>Parietal (clusters)</td>
<td>P7</td>
<td>POz</td>
<td>P8</td>
</tr>
<tr>
<td>Occipital (single electrodes)</td>
<td>O1</td>
<td>Oz</td>
<td>O2</td>
</tr>
</tbody>
</table>

Figure 4.2 A bird’s eye view of the scalp map with electrodes of interests.
Table 4.2

N170 and POz Positivity

Bars display the amplitudes for all the electrodes (clusters) corresponding to Table 4.1. Error bars show SEM.
4.3.1 ERP Findings

4.3.1.1 Model Differences

As I have only two computer-generated face avatar models, my conclusion is a conjecture to fulfill due diligent of the research exploration. In the current analysis, I compared the two simulated face models, Leah and Xyza with photo faces. Because I was interested in the model differences, I collapsed across the conditions (emotion and intensity) for each model and used the average in the analysis. ERP components of interest were the N170, POz positivity and P200. Post-hoc comparisons were adjusted with Bonferroni corrections.

**N170 Component.** I ran a 2x3 repeated-measures with electrode (P7 and P8 clusters) and model (Leah, Xyza, and photo) as factors. I observed a main effect of electrode, $F(1, 53) = 4.236, p = 0.045, \eta^2_p = 0.074$, which N170 amplitude is significantly larger at the right electrode cluster (-0.215 µV vs. 0.008 µV). There was also a main effect of model, $F(2, 106) = 11.919, p < 0.001, \eta^2_p = 0.184$. Pair-wise comparisons revealed that photo (-0.222 µV) evoked larger N170 responses than both Xyza (-0.075 µV) and Leah (-0.015 µV) ($ps < 0.001$), while there was no systematic differences of N170 amplitude between Xyza and Leah ($p > 0.05$).

**POz Positivity.** I ran a repeated measures ANOVA with model (Leah, Xyza and photo) as factors and found model as a main effect, $F(2, 106) = 18.962, p < 0.001, \eta^2_p = 0.263$. Follow-up post-hoc comparison showed three models were significantly different from each other. Leah (0.752 µV) elicited the largest POz positivity than Xyza (0.507 µV) and photo (0.318 µV) (all $ps < 0.05$).

Figure 4.3 displayed scalp distributions and waveforms for both the N170 and POz positivity.
Figure 4.3  a) Scalp distributions for the N170 and POz Positivity components (120-160ms) of the grand average: from left to right are Leah, Xyza and Photo models. b) Waveforms from left to right represents grand averages of Leah, Xyza and Photo at the P7 (left), POz (middle) and P8 (right) clusters from posterior sites. The shade area on the waveforms reflects the N170 component (at P7 and P8) and POz positivity (at POz). c) Bars display the amplitude of the N170 and the POz Positivity for Leah, Xyza and photo. The curved brackets indicate significant main effects of model. *p < 0.05. Error bars show SEM.
**P200 component.** I ran a repeated-measures 3*3 ANOVA with electrode (O1, Oz, and O2) and model (Xyza, Leah, and Photo) as factors. Figure 4.4 depicts the topographic maps and ERP waveforms elicited by three models at occipital electrodes respectively. The results showed that electrode yielded significant differences ($F(2, 106) = 6.909, p = 0.006, \eta^2_p = 0.115$). Post hoc tests showed that the amplitude for the right hemisphere was the strongest at 2.532 $\mu$V, followed by the middle site at 2.197 $\mu$V and the left hemisphere at 2.181 $\mu$V. There were no differences between the left and middle sides, $p > .05$. I also found a strong main effect on model, $F(2, 106) = 31.804, p < 0.001, \eta^2_p = 0.375$). Post hoc tests revealed that photo triggered the biggest positivity followed by Leah and Xyza (2.620 $\mu$V vs. 2.306 $\mu$V vs. 1.989 $\mu$V), the amplitudes differences were significant among three models (all $p$s < 0.05). I also found a significant interaction between electrode and model, $F(4, 106) = 3.170, p = 0.029, \eta^2_p = 0.056$). Post-hoc comparison showed that all models differed significantly at all three electrodes, with photo faces evoked the largest P200 followed by Leah and Xyza respectively maximum at the right side (electrode O2) ($p$s < 0.05).
a) Leah, Xyza, Photo

b) Amplitude (µV) vs Time (ms)

- Leah
- Xyza
- Photo
Figure 4.4 a) Scalp distributions for the P200 (170-280ms) for the grand averages of three models: from left to right are Leah, Xyza and photo. b) Waveforms from left to right represent grand averages of Leah, Xyza and photo at the O1 (left), Oz (middle), and O2 (right) clusters from occipital sites. The shaded area on the waveforms reflects the P200 component. c) Bars display the amplitude of the P200 for Leah, Xyza and photo over occipital sites: O1 (left), Oz (middle), and O2 (right). The curved brackets indicate significant main effects of site and model. *p < 0.05. Error bars show SEM.

4.3.1.2 Model Differences and Emotion & Intensity

The previous results demonstrated the three models (Leah, Xyza, and Photo) differed between the N170, POz, and P200 components. Thus, the current analysis explored whether emotion and emotional intensities affect individuals’ perception of realism for avatars. ERP components of interest remain unchanged from my previous analyses. Pairwise comparisons were adjusted with Bonferroni corrections.

N170. I ran a 2x3x2x3 repeated measures ANOVA with electrode (P7 and P8 clusters), model (Leah, Xyza, and Photo), emotion (anger and happiness) and intensity [neutral (0%), 20%, and 40%] as factors. The result showed only a main effect of model $F(2, 106) = 13.144,$
\[ p < 0.001, \eta^2_p = 0.199, \] where the N170 responded to photos much larger than Leah and Xyza (-0.223 \mu V vs. -0.080 \mu V vs. -0.014 \mu V). Post-hoc analysis also confirmed a larger N170 amplitude for photo than avatars (\( ps < 0.05 \)), while the two avatars depicted no systematic differences for the N170 (\( p > 0.05 \)). Neither main effects of emotion and intensity nor interactions between the factors were found.

**POz Positivity.** A 3x2x3 repeated measures ANOVA with model (Leah, Xyza, and Photo), emotion (anger and happiness) and intensity [neutral (0%), 20%, and 40%] as factors were conducted. Again, only model yielded a main effect, \( F(2, 106) = 21.736, p < 0.001, \eta^2_p = 0.291 \). Leah elicited a significantly larger POz amplitude, followed by Xyza and photo (0.752 \mu V vs. 0.497 \mu V vs. 0.307 \mu V). Pairwise comparison suggested these models distinguished from each other (\( ps < 0.05 \)).

**P200.** For this component, I ran a 3x3x2x3 repeated measures ANOVA with electrode (O1, Oz, and O2), model (Leah, Xyza, and Photo), emotion (anger and happiness) and intensity [neutral (0%), 20%, and 40%] as factors. I found main effects of electrode, \( F(2, 106) = 7.178, p = 0.005, \eta^2_p = 0.119 \), model \( F(2,106) = 35.228, p < 0.001, \eta^2_p = 0.399 \), and emotion, \( F(1, 53) = 5.336, p = 0.025, \eta^2_p = 0.091 \). The laterality showed a larger P200 at O2 in comparison to O1 and Oz (2.538 \mu V vs. 2.177 \mu V vs. 2.193 \mu V), which demonstrated the right hemisphere dominance. Post-hoc analysis also revealed no systematic difference of P200 amplitude between O1 and Oz (\( p > 0.05 \)). The strongest P200 response were found in photo faces (2.619 \mu V), followed by Leah (2.305 \mu V) and Xyza (1.984 \mu V). All corresponding pairwise comparisons were significant (\( ps < 0.05 \)), showing that the models differed from each other. For emotion, anger (2.336 \mu V) also elicited a larger P200 than happiness (2.269 \mu V).
A three-way interaction between electrode, emotion, and intensity was found, $F(4, 212) = 3.220, p = 0.027, \eta^2_p = 0.057$. Since I aimed to investigate if emotions or intensities projected effects on perceived model differences, interactions without model factors would not be analyzed further. No other interactions were returned by the ANOVA.

Taken together, I found no enough evidence to support emotional and intensity effects on the perception of model differences in realism.
4.3.2 Self-report Questionnaire Result

4.3.2.1 Human Likeness

I employed the same analysis method using a repeated measures ANOVA with models (Leah, Xyza and Photo), emotions (anger and happiness) and intensities (neutral (0%), 20%, and 40%) as factors. In this analysis, I focused on whether emotion and intensity interacted with the different models in the human-likeness score. I found a main effect on intensities, $F(2, 104) = 11.732, p < 0.001, \eta^2_p = 0.184$. Pairwise comparisons with a Bonferroni correction revealed that 40% expressions were viewed as more human than 20% and neutral expressions ($p < 0.05$) (7.086 vs. 6.786 vs. 6.623), while there was no difference between 20% and neutral expressions ($p > 0.05$). There was no main effect of model, but I found a significant interaction between model and emotion, $F(2, 104) = 4.486, p = 0.018, \eta^2_p = 0.079$. For both avatar models, happiness scored higher in human-likeness than anger, whereas the pattern was opposite in photo faces where anger was perceived more in human-likeness than happiness (see Figure 4.5).

![Figure 4.5](image)

*Figure 4.5 Bars represented the human-likeness score of Leah, Xyza and Photo for anger and happiness with three intensity levels (0%, 20%, and 40%). The curved brackets indicate significant main effect of intensity. *p < 0.05. Error bars show SEM.*
4.3.2.2 Correct Responses

I observed main effects of model, $F(2, 104) = 21.552, p < 0.001, \eta^2_p = 0.293$, emotion, $F(1, 52) = 49.584, p < 0.001, \eta^2_p = 0.488$, and intensity, $F(2, 104) = 25.922, p < 0.001, \eta^2_p = 0.333$. Photos had the highest correct responses, then Xyza, and Leah with the lowest correct responses (45.6% vs. 35.8% vs. 24.8%). All pairwise comparisons with a Bonferroni adjustment were significant ($ps < 0.05$). Happiness also resulted in more correct responses than anger (42.8% vs. 28.0%). Further, 40% expressions obtained a higher rate of correct responses than 20% and neutral (0%) expressions (51.9% vs. 22.2% vs. 32.2%). Post-hoc analysis with a Bonferroni adjustment revealed significant differences between 40% and the other two expressional intensities ($ps < 0.05$) and no systematic difference was observed between 20% and neutral (0%) expressions ($p > 0.05$). These effects are shown in Figure 4.6.

There was a three-way interaction of model, emotion, and intensity, $F(4, 208) = 6.313, p < 0.001, \eta^2_p = 0.108$. I broke it down into three two-way interactions based on model for further analysis. For Leah, there were main effects of emotions, $F(1, 52) = 12.042, p = 0.001, \eta^2_p = 0.188$ and intensities, $F(2, 104) = 17.065, p < 0.001, \eta^2_p = 0.247$, and a significant interaction of both factors, $F(2, 104) = 6.912, p = 0.005, \eta^2_p = 0.117$. Participants were more likely to be correct for expressions of happiness compared to anger (30.2% vs. 19.5%). Moreover, 40% expressions outscored 20% and 0% expressions for the correct responses (46.2% vs. 9.4% vs. 18.9%). 40% happiness seemed to be the easiest to recognize with 60.4% correct response rate, while people less likely recognized 20% anger with only 7.5% correct response rate.

For Xyza, there were also main effects of emotion, $F(1, 52) = 55.777, p < 0.001, \eta^2_p = 0.518$ and intensity, $F(2, 104) = 6.716, p = 0.004, \eta^2_p = 0.114$, and a significant interaction of both factors, $F(2, 104) = 38.247, p < 0.001, \eta^2_p = 0.424$. Similar to Leah, participants obtained significantly higher correct response rate for happiness than anger (50.9% vs. 20.8%).
40% and 0% expressions generated higher correct response rate than 20% (41.5% vs. 20.8% vs. 45.3%), while pairwise comparisons with a Bonferroni correction showed no systematic difference between 40% and 0% expressions ($p > 0.05$). Again like Leah, 40% happiness had the highest correct response rate, whereas 20% has the lowest (77.4% vs. 7.5%).

For photo, I found a main effect of intensity, $F(2, 104) = 19.126$, $p < 0.001$, $\eta^2_p = 0.269$, and a significant interaction between emotion and intensity, $F(2, 104) = 27.601$, $p < 0.001$, $\eta^2_p = 0.347$. People recognized 40% expressions better than 20% and 0% expressions (64.2% vs. 36.3% vs. 36.3%). Furthermore, 40% happiness obtained the highest correct response rate while 20% happiness has the lowest (76.4% vs. 29.2%).

To summarize, happiness contributed to increased human-likeness for avatar faces overall, but did not affect human-likeness between the two avatar models. Responses were more likely to be correct for expressions of happiness compared to anger. Xyza had the highest correct responses in happiness amongst three models. Taken together, I have some evidence suggests that happiness makes Xyza more human and recognizable compared to Leah in the perceptual tasks.
Figure 4.6 Bars represent the correct response rate of Leah, Xyza and Photo for anger and happiness with three intensity levels (0%, 20% and 40%). The curved brackets indicate significant main effects of model, emotion and intensity. *p < 0.05. Error bars show
4.4 Discussion and Assumption

Two avatars used in my experimental paradigm differed from one other on multiple dimensions, the most salient of which were realism, attractiveness, and apparent age. The purpose of this exploratory analysis was to examine whether and how the ERP and behavioral responses to the two avatar models differed. These analyses were necessarily exploratory, since any of the relevant dimensions or interactions between them may have contributed to the differences found in my data. To my knowledge, there are not yet any published reports in the literature investigating the ERP responses to differing levels of realism in human agents.

In the survey instrument, the photo and two avatar faces generated no significant differences in the human likeness rating, although I found that 40% intensity expressions contributed to increasing ratings of realism, and happiness also tended to make the avatars look more human. Furthermore, participants correctly recognized the emotion on photo faces than on the two avatar faces. Between the avatar faces, participants responded more correctly to Xyza than Leah. Participants identified 40% expressions and happiness the most, while they experienced difficulties in identifying 20% expressions, in particular with 20% anger.

Positive expressions, like happiness, enhanced the realism of the avatars. This enhancement is in line with a happy-face recognition advantage where people are faster and more accurate to recognize happy expressions over neutral and negative expressions, even for schematic faces (Kirita & Endo, 1995; Leppanen & Hietanen, 2004). Two explanations may underlie this “happy-face advantage”. Firstly, configurally, happiness contains more physical changes and salient features than negative expressions (Leppanen & Hietanen, 2004). One can identify and differentiate happiness on the basis of only a few (or sometimes even a single) features, such as a pulled-up mouth corner or wrinkles around the eyes (Adolphs, 2002). Therefore, studies report that people are less able to distinguish between negative and neutral
expressions compared to between happy and neutral expressions (Johnston, Katsikitis, & Carr, 2001). Secondly, Ohman et al. (2001) asserted that happiness, compared to negative expressions, is easier to “pose on demand”. This is an essential point to our avatar models as they were created based on both posed and spontaneous facial expressions from real actresses. As happiness can be posed more naturally compared to anger, virtual agents displaying happy expressions may therefore be perceived as more human-like.

In general, I found that the POz positivity and the P200 appeared to be sensitive to between-avatar differences, while the N170 reponded to the differences between photo and avatar faces. There are some studies reporting that there are distinct neural ERP signatures that differentiate attractive versus unattractive faces (Halit et al., 2000) and young versus old adult faces (Schweinberger et al., 2010). More details regarding how attractiveness and age modulate the ERP components will be discussed in the next session (4.4.3). Therefore, if ERPs in this analysis do not resemble these ERP differences in age or attractiveness, in terms of spatial or temporal distribution in existing literature, then the observed ERP differences may be attributable to realism alone – or at least to some other aspect of the images. As my findings were predicated on the only two models available at the time when the experiment was conducted, I hope to use the ERP effects observed here to generate hypotheses for future studies when more avatar models become available.

4.4.1 N170 and POz Positivity

I found that both the N170 and the POz positivity were modulated by the differences between the two avatar models and photo faces. The N170 turned out to differentiate between human and avatar faces rather than between the individual avatar faces, while the POz positivity responded differently to each of the three conditions. I did not find effects of either emotion or intensity main effects on the N170 or POz positivity.
The present analysis of N170 reprises my findings from Chapter 3, in which avatar faces attenuated the N170 amplitude compared to photos. The N170 was not significantly different between Xyza and Leah. This lack of a difference between the models could suggest that the N170 is not sensitive to more fine-grained levels of realism between different virtual models. At this level, a more general differentiation between real and non-real face occurs with the N170. Gajewski and Stoerig (2011) raised a question of whether faces of different species that share spatially similar features are processed categorically or prototypically. They presented participants with stimuli of varying categories, including human, monkey, and dog faces, with doors as non-face control stimuli. They found that the amplitude of the N170 was attenuated when stimuli deviate away from human categories, such that human faces evoked the highest N170 amplitude, followed by monkeys, then dogs, and finally doors. They concluded that the N170 displayed a preference for the categorical processing of faces.

Likewise, Campbell, Pascalis, Coleman, Wallace and Benson (1997) investigated the species boundaries of face processing by using morphed faces of humans, monkeys and cows. In their study, participants were instructed to perform a force-choice discrimination task to indicate which category (human, monkey, or cow) that the morphed face belonged to. Participants produced well-discriminated species categories for monkey-cow and cow-human faces, but not human-monkey faces. This finding also supported the categorical processing of humanlike faces. Studies involving people with prosopagnosia (face blindness) have showed that prosopagnosics are more resilient to animal faces, a finding which also supports the categorical processing of humanlike faces. In other words, some prosopagnosics were able to recognize animal faces, such as sheep (McNeil & Warrington, 1993), cows, and dogs (Bruyer et al., 1983), but their recognition of human faces was specifically impaired. Assal et al. (1984; as cited in Campbell et al., 1997) also observed that the relearning of human faces was required after recovering from prosopagnosia, while this relearning process was
unneeded for animal faces. Taken together, the impairment of recognizing the specific
category of faces and separate learning processes between human and animal faces for
recovering from prosopagnosia patients lend further support to a categorical account of facial
processing.

Gajewski and Stoerig (2011) asserted that their results favored categorical processing of faces
because the N170 evoked by other species was attenuated compared to human faces. This
agrees with my findings that virtual and human faces were perceived as belonging to distinct
categories, which resulted in N170 differences. As a result, the indiscrimination between
Xyza and Leah may be attributable to the N170’s insensitivity to the same-category
differentiation. On the other hand, the amplitude differences between Leah and Xyza at the
POz positivity may indicate some within-category distinction, perhaps due to prototypical
processing.

I postulate that the two processing strategies (categorical and prototypical) may be
interchangeable and integrable. This is especially true when it comes to hyper-realistic virtual
agents that sit on the boundary between categories (Cheetham et al., 2011). To be more
specific, when two faces are on opposite sides of a clear categorical boundary (e.g. human
versus avatar faces), our brain can easily process the faces and determine which category is
more human. When two faces are categorically indistinct (e.g. between two avatar faces),
prototypical processing is activated as a secondary resource to further discriminate the faces.
This may be evident by my findings that the two avatar models displayed no differences in
the N170 amplitude, while within the same time window, the POz positivity was
distinguished between the two models. Tentatively, this might indicate that the POz positivity
could be an indicator to identify different degrees of realism in virtual agents. Of course, the
possibility remains that the lack of avatar models’ differences for the N170 may be
influenced by overlap with the generator(s) of the POz positivity, or vice versa.
To confirm the POz positivity as an indicator to human realism, we need to carry out more studies. Extrapolating from my data, if POz positivity exists as a strong indicator, this opens up for a new whole set of questions:

a) Do we have a specific region of the brain and/or an ERP component that is sensitive to different degrees of realism in avatar faces? Alternatively, do realism levels across avatars actually matter, given that our brain can differentiate avatar faces from human faces?

b) Could the POz positivity be a psychophysiological indicator of the uncanny valley, of an avatar looking exceptionally similar to a human? I found a centrally distributed area of positivity at the posterior-occipital site on the topographic maps that was prominent for both avatar models. Furthermore, the findings showed that the POz particularized response between avatar models, while the lateral sites (P7 and P8) responded to only photo-avatar differences. This leads to my next question.

c) If the POz and N170 are inter-related and co-processed, when are virtual faces involved in early processing? In other words, when processing avatar faces, one pathway (N170) identifies deviations from an average human face, whereas a parallel pathway (POz positivity) identifies differences between virtual faces.

4.4.2 P200

In support of the hypothesis, I found that the P200 was largest in amplitude for photo faces, followed by Leah and then Xyza,. As I discussed in the previous chapter (Chapter 3), the P200 has been associated with face typicality (Halit et al., 2000; Stahl et al., 2008). Therefore, I can surmise that participants perceived Leah’s face to be more typical of a human face than Xyza. This is consistent with the participants’ perception of Leah appearing more real and alive than Xyza. The attenuation of the P200 from photo to Leah to Xyza illustrated that, despite the strong realism of the Leah model in particular, photographic
images still resembled the average human face more than both avatars, while Leah resembled
the average human face more than Xyza.

4.4.3 Influences of Attractiveness and Age
It is worth reiterating that my conclusions based on ERP responses are necessarily tentative
because there are a number of physical differences between the two avatars that must be
taken into account. Informal inspection of the two avatar models reveals that Leah appeared
to be older and less conventionally attractive than Xyza.

The N170 has been shown to generally respond to facial attractiveness stimuli, but there is
great inconsistency in this across studies. For instance, some studies have found a reduced
N170 amplitude for attractive faces compared to less attractive faces (Halit et al., 2000;
Trujillo, Jankowitsch, & Langlois, 2014), whereas others have reported enhanced N170 to
attractive faces (Hahn et al., 2016; Zhang & Deng, 2012). These directional differences
could be explained by a number of factors, including varying identities, low-level stimulus
differences, or other methodological differences between studies. Nevertheless, the
processing of attractiveness at an early stage is evident in all these studies. In my analysis,
N170 amplitudes differed between photos and avatars, but not between the individual avatars.
I am not inclined to interpret the absence of N170 amplitude difference between avatars as
evidence that Leah and Xyza were perceived equally attractive/unattractive. Rather, a more
plausible explanation may be attributable to realism. The brain devotes primary cognitive
resources to distinguish human from non-human faces, and processing of other, within-
category dimensions (e.g. attractiveness or age) becomes secondary and is processed only
when there are sufficient spare resources or appropriate task demands.

A robust finding on facial attractiveness from ERPs is the modulation of the late positive
complex (LPC), a component emerging between 400 and 600ms after the stimulus onset.
Attractive faces evoke an enhanced LPC (Johnston & Oliver-Rodriguez, 1997). Additionally, Lu, Wang, Wang, Wang and Qin (2014) investigated how facial attractiveness modulated the ERP response using cartoon faces and found a larger LPC for attractive cartoon faces. There are also some studies that report early ERP modulations for facial attractiveness. Werheid, Schacht and Sommer’s (2007) study indicated that the processing of attractiveness can be initiated within 300ms. They found an enhanced early posterior negativity (EPN) between 230-280ms to attractive faces. Likewise, Chen et al. (2012) found attractive faces evoked a smaller P2 (130-170ms) and a larger N2 (180-230ms) at frontal and frontal-central sites. Rellecke, Bakirtas, Sommer and Schacht (2011) reported a sustained larger posterior negativity from 260 to 1000ms for attractive faces. Compared to my findings of the POz positivity and P200, these earlier components in existing literature have different temporal and spatial distributions. The temporal distribution of the POz positivity overlaps with P2 in Chen et al.’s (2012) study, but the spatial distribution is different. POz in my study maximized at the parietal-occipital site, while Chen et al.’s (2012) P2 maximized at the frontal site. Since the two components were both positive, I may not consider they are the dipolar counterparts of each other. For the P200, the temporal comparable components in the facial attraction literature is the EPN (Werheid et al., 2007). Yet there is no evidence suggesting the P200 and EPN share similar responses to face attractiveness. Collectively, I have some evidence indicating that different attractiveness levels between the two models may not contribute to the the ERP differences found for the POz positivity and P200 in my study.

Regarding the age dimension, Leah is actually and perceptually older than Xyza. According to Schweinberger et al. (2010), humans are experts in capturing age-related features on a face including configural changes, changes in skin texture, and color. An “own-age bias” (OAB) is often observed in face processing, as changes in facial features ultimately influence face
recognition memory depending on observers’ own age. Some studies (Fulton & Bartlett, 1991; He, Ebner, & Johnson, 2011; Wolff, Wiese, & Schweinberger, 2012) showed that young participants were particularly discriminative in recognizing young faces, while elderly participants did not present the OAB effect. Other studies (Anastasi & Rhodes, 2005), observed the opposite—elderly participants demonstrated OAB and young participants did not. Such inconsistencies in the OAB findings can derive from 1) participants’ age, especially young participants’ age ranging between 3 months to 5 years (Rhodes & Anastasi, 2012); and 2) the potential impacts of the control of different facial features. Besides, a number of ERP studies examined the interaction between the OAB and facial recognition memory between young and elderly adults. Elevated N170 and attenuated posterior P200 amplitudes were reported for old face stimuli and more pronounced for young participants (Wiese et al., 2008). An increased N250 amplitude was also observed for young faces in young participants, reflecting the OAB effect (Wiese et al., 2008; Wiese, Kachel, & Schweinberger, 2013). On the contrary, I observed an enhanced P200 for Leah compared to Xyza, which if age-related facial characteristics were driving the effect, ought to be reversed given that most of the participants’ age are closer to Xyza. Additionally, Wiese, Kachel and Schweinberger (2013) also argued that the N170 and P200 may not account for the OAB in recognition memory, since some evidence suggests that N170 and P200 amplitudes’ alternation were also present in older participants (Wiese, Komes, & Schweinberger, 2012) who failed to show bias to own-age faces. Again, inconclusive evidence does not argue convincingly that the ERP differences in my results were the responses to age-related facial features’ differences between the two models.

Additionally, there is considerable limitation in directly comparing my findings to previous literature regarding facial attractiveness and age effects, because there is lack of studies of comparable ERP components. I can only compare reported components in the literature that
fall within the similar temporal distribution but not necessarily in the same spatial
distribution, due to some of the components being maximal at different electrodes locations.
As such, I cannot rule out the idea that realism levels are the major factor driving the ERP
differences in my findings.

4.4.4 Superstimuli

With my observation during the experiment, participants were happy to endorse Leah as
human whereas they commented that Xyza appeared airbrushed and somewhat
“overproduced” (e.g. smoothing for flaws to achieve perfect skin look). It is important to
emphasize that Xyza was not intended to be unrealistic. Rather, Xyza was primarily designed
to be attractive to a wide range of users, and it is possible that this attractiveness comes at the
expense of realism via enhancing features to disguise flaws. In other words, Xyza may in
effect be a “superstimulus” (also known as supernormal stimulus). This concept was initially
proposed by Niko Tinbergen in 1948. Superstimuli or supernormal stimuli are stimuli with
features that are exaggerated, and which will evoke an enhanced behavioral response
compared to the original stimuli (Tinbergen, 1948). He found an interesting phenomenon in
animals like insects or fish, who would prefer an artificial object with exaggerated features
that are biologically attracting members of the same species. He studied a type of fish named
the stickleback, in which the characteristic of a bright red underbelly in the males acted as a
territorial sign or a mating sign for female sticklebacks in breeding seasons. Tinbergen
prepared two dummy models (model R) with redness on the ventral side: one represented
dead sticklebacks and the other one represented schematic models, and a control model
(model N), which consisted of neutral colored males in the non-breeding season and females.
He was interested in testing if both the males and females would target (either attacking or
being attracted) to the dummies. Intriguingly, both males and females were attracted by
model R. Aside from sticklebacks, Tinbergen also found other species like birds and lizards,
would target an artificial dummy rather than a real and biologically significant subject. Lorenz (1981) termed these stimuli “releasers”, while Tinbergen coined it “sign stimuli”. Furthermore, Lorenz (1981) highlighted that sign stimuli dependence only happened in innate behaviors, which suggested that learning was in the absence of the responses to these stimuli. Such a behavioral process was also named as the innate releasing mechanism (Lorenz, 1981).

According to Lorenz (1981), individuals perceived kittens, puppies, cartoon characters, or babies as cute because of their disproportional features of the face, such as bigger eyes, a smaller nose, and a larger head (Barrett, 2010). McKelvie’s (1993) study showed that larger eyes and lower facial feature placement of a schematic baby face (Face 1, see Figure 4.7) were perceived as cuter. In the same vein, individuals are attracted to cartoon faces because of the disproportional or exaggerated facial features. For animators, such exaggeration aims to magnify facial expressions and emotional engagement (Thomas & Johnston, 1995).

Interestingly, in our daily life, people try to attract others by enhancing the “releasers”. A good example would be facial cosmetics which are used to create “visual illusions” by emphasizing particular features of the face. For instance, eyes are visually enlarged using different eye shadowing techniques, or artificial/extended eyelashes, and pimples or face spots are concealed using concealers and foundations to give an illusion of clear and perfect skin. These techniques are dedicated to exaggerating these “releasers” to generate preference from others compared to a person’s natural status.
Figure 4.7 Schematic baby faces in McKelvie’s (1993) study. Face 1 has larger eyes and low feature placement. Face 2 has the same feature placement with Face 1 except for smaller eyes. Face 3 moves features in Face 1 upwards. Face 4 has the same feature placement with Face 3 except for smaller eyes.

The presence of superstimulus-like features in Xyza has raised a question regarding the trade-off between attractiveness and realism when designing human agents. With the purpose of minimizing the possibilities that people emotionally reject avatars, engineers or computer graphic designers tend to weigh attractiveness over realism and turn avatars into superstimuli to some extent. Computerized 3D avatars applied in commercial fields or movies appear to be airbrushed with overly smoothed skin, disproportional facial features and somewhat exaggerated movements to appeal to audiences or viewers.
4.4.5 Next Steps

To further address the questions of the current analysis, I propose some future steps for this program of research.

Firstly, incorporating different avatars with a range of realism levels to confirm or disconfirm the POz positivity and its relationship with the N170. For instance, studies can incorporate cartoonish avatars, realistic avatars (such as FaceGen) and hyper-realistic avatars (the AFS) to see whether the POz positivity (if present) and the N170 will be modulated by the different realism levels.

Secondly, controlled experiments using avatar stimuli with a range of age, attractiveness and realism levels can further investigate if age or attractiveness influence the realism levels of avatars. Future studies can have participants rate the humanness of avatars with different age groups (e.g. younger group - 20-25 years old; middle aged group – 35-40 years old; and elder group – 60-65 years old) or different attractive levels. Researchers can investigate whether ERP responses correspond to the humanness rating with age or attractiveness as independent variables.

Thirdly, it is important to create a more validated survey or even interview to gather participants’ perception over the avatars. Surveys may include items that represent the human likeness and affinity dimension of Mori’s uncanny valley, and tested for validity and reliability. This could further aid the interpretation of ERP results.

4.4.6 Conclusion

The findings and conclusions are necessarily tentative because the two models used in my experiment had different characteristics, producing different behavioral and neural responses. Nevertheless, the analysis and results from this unplanned analysis shed light on hypotheses for further research into the realism exhibited in avatars. My initial hypotheses were not
designed to probe the effect of attractiveness or age-related differences in the facial features of the two human agents. Rather, I initiated the exploratory analysis because I was driven by the curiosity of if and how hyper-realistic avatars can differ and whether these differences can be disclosed at a psychophysiological level.

Even when using the same platform and computation, my findings indicated that the avatars still differed in realism. Accordingly, designing avatars or artificial intelligent agents should go beyond the computational process and incorporate multiple dimensions to include emotions and facial expressions. Furthermore, avatars are not merely the production of sophisticated computational platforms. At a more fine-tuned level, subtle changes in the appearance can result in realism differences between avatars.

The findings also help to extend the field where we can use avatars to design more systematic and controllable experiments. When avatars have strong physical similarity, paired with identical psychophysiological response to real-life photos, we can obtain greater control than photos. For instance, we can take the same identity that varies in attractiveness or age-related facial characteristics in an avatar, which is much harder to do in a photo. Therefore, by utilizing avatars in experiments interested in investigating human expression, researchers can be more systematic in exploring the effects of age and attractiveness, while controlling variables like identity. The manipulation in avatars can be more fine-tuned and accurate than photographs. While this is promising implementation in the future, it is not a realizable goal right now with the current computer technology.
Chapter 5 Group Differences

5.1 Introduction

In previous chapters, I explored how our brain responded to humanlike avatar faces (Chapter 3) and whether the differences between two existing avatars contributed to realism differences (Chapter 4). The results in Chapter 4 revealed that two avatars differed in realism despite their being developed using essentially the same methodology. Given that I only had two complete avatars to work with, the ability to interpret differences in response was necessarily limited. I think it would worthwhile to further the analysis from the observers’ perspective as they anthropomorphize human agents with various levels of realism or “humanness”. Namely, if a viewer perceives a human avatar as having a higher degree of realism than another viewer, will the brain activity mirror the perceptual differences between the viewers? More specifically, are there any differences in ERP responses between those who see simulated faces are more human and those who see simulated faces less human?

In social cognition, Kunda (1999) argued that it is a basic human cognitive function to anthropomorphize broad classes of objects, including animals, robots or human agents. We tend to imbue nonhuman agents with human features and mental capabilities. Yet, there are few studies of individual differences in the tendency to anthropomorphize, or in the degree to which people perceive humanlike traits in artificial or virtual agents. Nowak and Rauh (2006) had participants evaluate a number of avatars and found participants tend to use more anthropomorphic avatars to represent themselves in virtual environments. Waytz, Cacioppo and Epley (2010) proposed the IDAQ (individual differences in anthropomorphism questionnaire) and concluded that individual differences in anthropomorphism could predict the degree of moral care and concern, responsibility and trust, and a source of social influence projected to virtual agents. At the neuroimaging level, Cullen, Kanai, Bahrami and
Rees (2014) used MRI to investigate the neural basis of anthropomorphism. They observed a correlation between anthropomorphic attribution – the tendency to perceive humanlike attributes in nonhuman stimuli – and grey matter density of the left temporoparietal junction, which is an area that has been implicated in mentalizing. Despite these few efforts to look at individual differences in perceiving human likeness, there is not yet an emerging picture of what drives these differences, or what consequences there will have for the use of virtual agents.

In spite of some scant evidence of individual differences in perceiving anthropomorphic features of non-human objects, we know very little about how such perception of anthropomorphism or realism can alter people’s responses to non-human agents, not to mention hyper-realistic avatars that somehow blur the boundary between human and virtual agents. Understanding individual differences in responses to avatars is not only important for identifying perceivers who are more likely to accept virtual agents as human, but also for understanding why some people are more inclined to reject such agents as humanlike.

In this chapter, I explore individual differences in the response to avatars and photo images based on their responses to the survey questionnaire. Given that I have an adequate number of participants to allow me to group them according to their ratings of the two avatars (Xyza and Leah), I was able to test if perception of realism is reflected in the ERP components I have identified in prior chapters. Based on my previous findings, the N170 (the face-sensitive component; Bentin et al., 2007) and the P200 (sensitive to face typicality; Wuttke and Schweinberger, 2019) seem to be most likely to be fruitful in exploring the following hypotheses:

**H5.1** People who treat simulated faces more real than photo faces will show larger N170 and P200 amplitudes to avatar faces compared to those who see simulated faces as less real.
H5.2 Those who perceive both simulated and photo faces equally as highly humanlike will show larger N170 and P200 amplitudes than those who rate both simulated and photo faces as less humanlike.

H5.3 Individuals rating high on the human likeness dimension for simulated faces will have larger N170 and P200 amplitudes compared to those who rate simulated faces lower on this dimension.
5.2 Methods and Results

In the behavioral study (reported in Chapter 3), I found that there was no significant distinction between photo and simulated faces in the human likeness score. This is against my initial expectation as the majority of people tended to the two models as equally human - or equally not human. Considering the prominent differences found between photo and simulated faces in terms of ERP amplitude in Chapter 3, I wondered whether the differences in the perception of humanness between the two avatars could explain the incongruent findings between questionnaire and ERP in Chapter 4. Accordingly, I investigated whether those who perceive simulated faces as more human will have a different ERP response than those who see simulated faces as less human.

I categorized the participants into three groups, using three different methods based on their human likeness score in the questionnaire.

The three methods are:

1) **Difference score**: a difference score derived by subtracting the human likeness ratings for photo faces from those for simulated faces. The larger the difference score value is, the more humanlike the avatar faces are rated. As a result, I named the first group which contains first 13 participants with the largest difference scores “Simulation More Real (SMR)” group; the middle group of 13 participants with different scores close to “0” I named “Simulation Equals Photo (SEP)”; and the third group of 13 participants with the largest negative difference-score values I named the “Simulation Less Real (SLR)” group.

2) **Average score**: average human likeness scores for simulated and photo faces. Once the average score was computed, I selected the first and last 13 participants and named the high average score group “High Humanness (HH)” and the low average
score group “Low Humanness (LH)”. The middle group of 13 participants was
selected with scores closest to the median, and I named the group “Medium
Humanness (MH)”.

3) Simulation score: grouping based on human likeness ratings for simulated faces only.
The first group “High Simulation (HS)” consists of 13 participants who rated the
highest scores of simulated faces and the second group “Low Simulation (LS)” has 13
participants who rated the lowest of human likeness scores on simulated faces. The
middle group “Medium Simulation (MS)” was identified with 13 participants whose
human likeness scores were closest to the median.

To sort participants into three groups based on each of these methods, I first ranked
participants from low to high on each method. I took the first and fourth quartile as extreme
groups, and compared them to a similar number of participants whose scores were in the
middle of the distribution. That is, I identified three groups – two extreme and one central –
but minimized any overlap between the groups by only considering quartile-sized subgroups.
This allowed me to focus more on scrutinizing ERP differences between extreme groups,
taking the middle group as a baseline. This approach gave 13 participants in each group, with
a further 15 participants (at the boundaries between groups) excluded from this analysis.

For each method and each group, I looked the amplitudes of the N170 and the P200
components. For each component, I took the average amplitude in the specified time
windows (same as in previous chapters) for each group.

For each of the three sorting methods, I performed a repeated-measures ANOVA with groups
(3 levels), realism (simulation and photo), and electrode [2 levels for the N170 (P7 and P8
cluster), and 3 levels for the P200 (O1, Oz and O2), with Bonferroni corrections applied. To
avoid redundency, this analysis did not include emotion or intensity as factors.
5.2.1 Method 1: Grouping based on difference score

**N170 component.** The ANOVA revealed main effects of both group, \( F(2, 24) = 5.822, p = 0.019, \eta^2_p = 0.327, \) and realism, \( F(1, 12) = 15.535, p = 0.002, \eta^2_p = 0.564. \) There was no significant main effect of electrode. Pair-wise comparisons for groups showed larger negativity of the N170 for the medium group SEP (\(-0.337 \, \mu V\)) relative to SMR (\(-0.076 \, \mu V\)) and SLR (\(0.297 \, \mu V\)) (\(p < 0.05\)), and SMR evoked a significantly larger N170 than SLR (\(p < 0.05\)). In terms of realism, photo faces (\(-0.131 \, \mu V\)) displayed a larger N170 amplitude than simulated faces (\(0.054\)). Scalp distributions, waveforms, and bar graph depicting these effects are shown in Figure 5.1.

![Scalp distributions, waveforms, and bar graph depicting these effects](image)

**Figure 5.1**: a) Scalp distributions, waveforms, and bar graph depicting these effects.
Figure 5.1 a) Scalp distributions for the N170 (120-160ms) for the grand averages of three groups: from left to right are SLR, SEP and SMR. b) Waveforms from left to right represent grand averages of the three groups (SLR, SEP and SMR) at the P7 (left) and P8 (right) clusters from posterior sites. The shaded area on the waveforms reflects the N170 component. c) Bars display the amplitude of the N170 for groups (SLR, SEP and SMR), and realism (photo and simulation). The curved brackets indicate significant main effects of group and realism. *p < 0.05. Error bars show SEM.

**P200 component.** The ANOVA revealed a main effect of realism, $F(1, 12) = 36.610, p < 0.001, \eta^2_p = 0.753$. The amplitude of P200 for photo faces (2.661 µV) was significantly larger than simulated ones (2.222 µV). While the SLR group (2.761 µV) yielded the largest P200 amplitude, followed by SMR (2.557 µV) and SEP (1.986 µV), there were no statistically significant effects involving group or electrode. The effects are displayed in Figure 5.2.
Figure 5.2  a) Scalp distributions for the P200 (170-280ms) for the grand averages of three groups: from left to right are SLR, SEP and SMR. B) Waveforms from left to right represent grand averages of the three groups (SLR, SEP and SMR) at the O1 (left), Oz (middle) and O2 (right) electrodes from occipital sites. The shaded area on the waveforms reflects the P200 component. c) Bars display the amplitude of the P200 for groups (SLR, SEP and SMR), and realism (photo and simulation). The curved brackets indicate a significant main effect of realism. *p < 0.05. Error bars show SEM.
5.2.2 Method 2: Grouping based on average score

**N170 component.** I found a main effect for realism, $F(1, 12) = 14.439, p = 0.003, \eta^2_p = 0.546$, with photo faces (-0.255 µV) having a larger N170 compared to simulated faces (-0.091 µV). I failed to observe a significant result for group, $F(2, 24) = 0.849, p = 0.440, \eta^2_p = 0.066$. HH (-0.387 µV) generally had the largest N170 amplitude, followed by MH (-0.117 µV) and LH (-0.016 µV). There were no significant effects involving electrode. These effects are depicted in Figure 5.3.
P200 component. There was no systematic difference in the P200 amplitude for group, $F(2, 24) = 0.744, p = 0.486, \eta^2_p = 0.058$, although the MH (3.037 µV) group showed larger P200 compared to both the HH (2.120 µV) and LH (2.244 µV) groups. Photo faces (2.688 µV) also exhibited significantly larger P200 amplitude than simulated faces (2.246 µV), $F(1, 12) = 30.909, p < 0.001, \eta^2_p = 0.720$. There was a main effect of electrode, $F(2, 24) = 9.590, p = 0.004, \eta^2_p = 0.444$, which further pair-wise comparison yielded that O2 (2.756 µV) had significantly larger P200 amplitude than O1 (2.289 µV) and Oz (2.357 µV) ($ps < 0.05$), while O1 and Oz showed no difference in P200 amplitude ($p > 0.05$). A significant interaction between group and electrode was obtained, $F(4, 48) = 3.584, p = 0.050, \eta^2_p = 0.230$, with MA group evoked significantly larger positivity at Oz than O2 ($p < 0.05$). These effects are depicted in Figure 5.4, below.
Figure 5.4 a) Scalp distributions for the P200 (170-280ms) for the grand averages of three groups: from left to right are LH, MH and HH. B) Waveforms from left to right represent grand averages of the three groups (LH, MH and HH) at the O1 (left), Oz (middle) and O2 (right) electrodes from occipital sites. The shaded area on the waveforms reflects the P200 component. c) Bars display the amplitude of the P200 for electrodes (O1, Oz and O2), groups (LH, MH and HH), and realism (photo and simulation). The curved brackets indicate significant main effects of electrode and realism. *p < 0.05. Error bars show SEM.

5.2.3 Method 3: Grouping based on human likeness score of simulation only

N170 component. The result yielded main effects of realism $F(1, 12) = 6.931, \ p = 0.022, \ \eta^2_p = 0.366$, with N170 amplitude being larger for photo faces (-0.222 µV) compared to simulated faces (-0.072 µV), and electrode, $F(1, 12) = 6.263, \ p = 0.028, \ \eta^2_p = 0.343$, with N170 being larger at the P8 cluster (-0.248 µV) than P7 cluster (-0.047 µV). The group effect failed to reach the threshold for significance, $F(2, 24) = 0.307, \ p = 0.702, \ \eta^2_p = 0.025$, which HS (-0.288 µV) generated the largest N170 amplitude, followed by LS (-0.100 µV) and MS (-0.054 µV). There were no significant interactions between these factors. Figure 5.5 shows the scalp distribution, waveforms and bar graphs of these effects.
a) LS, MS, HS

b) P7 Cluster, P8 Cluster

Amplitude (µV)

Time (ms)
Figure 5.5 a) Scalp distributions for the N170 (120-160ms) for the grand averages of three groups: from left to right are LS, MS and HS. B) Waveforms from left to right represent grand averages of the three groups (LS, MS and HS) at the P7 (left) and P8 (right) electrode clusters from posterior sites. The shaded area on the waveforms reflects the N170 component. c) Bars display the amplitude of the N170 for groups (LS, MS and HS), electrodes (P7 and P8 clusters) and realism (photo and simulation). The curved brackets indicate significant main effects of electrode and realism. *p < 0.05. Error bars show SEM.

P200 component. I found a main effect of realism, $F(1, 12) = 19.585, p = 0.001, \eta^2_p = 0.620$, such that P200 amplitude was larger for photo (2.325 µV) than simulated (1.896 µV) faces. I did not find a significant effect of group, $F(2, 24) = 0.440, p = 0.649, \eta^2_p = 0.035$. The MS (2.485 µV) group evoked a larger P200, followed by HS (2.025 µV) and LS (1.821 µV) groups. There were no significant interactions between these factors. These effects are shown in Figure 5.6.
Figure 5.6 a) Scalp distributions for the P200 (170-280ms) for the grand averages of three groups: from left to right are LS, MS and HS. B) Waveforms from left to right represent grand averages of the three groups (LS, MS and HS) at the O1 (left), Oz (middle) and O2 (right) electrodes from occipital sites. The shaded area on the waveforms reflects the P200 component. c) Bars display the amplitude of the P200 for groups (LS, MS and HS), and realism (photo and simulation). The curved brackets indicate a significant main effect of realism. *p < 0.05. Error bars show SEM.
5.3 Discussion and Assumption

I investigated whether differences in face-selective ERP components – specifically the N170 and P200 – would reflect differences between perceivers in terms of how they perceived avatar and photo faces. I aimed to explore whether individuals’ comparison between simulated and real photo faces and their perception of simulated face would be reflected in ERP differences. I have divided the participants based on three different methods, namely difference score, average score and simulation score. Overall, the findings did not support the idea of hypothesized that individual differences in realism perception would result in clear ERP signatures of this perception. I found that ERP differences between simulated and photo faces seemed to be independent of individual differences in perception.

Across the analysis, the only significant effect of group was in the difference score grouping method (Method 1), in which the SEP group elicited a larger N170 followed by SMR and SLR. In this analysis, I introduced the middle groups as a baseline to assist in interpreting the directions of ERP amplitude of the extreme groups, due to the lack of relevant literature as a reference. The result partially supported my first hypothesis, as I expected a magnified N170 amplitude for SMR over SLR. Participants who regard simulated faces more real than photo faces see both simulated and photo faces as more “facey”. Intriguingly, though, the SEP group has the largest N170, which was not my initial expectation – and somewhat undermines my conclusion of support for H5.1. There are a couple of possible interpretations that may explain the findings. The SEP group, in which people generally rate simulated faces are equally real or unreal to photo faces, may have a stronger tendency to anthropomorphize. This means that they are more accepting of face stimuli no matter human or non-human as faces. Studies have shown
that anthropomorphism has activated a number of regions, including the left temporoparietal junction (TPJ) and the anterior cingulate cortex (ACC) (Chaminade, Hodgins, & Kawato, 2007) and the posterior superior temporal sulcus (pSTS) (Heberlein, 2008). A study by Kuhn, Brick, Muller and Gallinat (2014) demonstrated a positive association with anthropomorphism score of cars and activities in the right fusiform face area (FFA), which is reported to be one of the neutral sources that generate the N170 (Horovitz, Rossion, Skudlarski, & Gore, 2004; Jacques et al., 2018; Yovel, 2016). It thus may be that the SEP anthropomorphized more than the other two groups, while the SMR group, with higher human likeness scores for simulated faces ascribed to anthropomorphism produced a larger N170. With the SLR group that simulated faces were regarded less human, the degree of anthropomorphism may be weaker and result in a smaller N170. This explanation, while intriguing, is clearly post-hoc. To explore it further, we can incorporate the IDAQ (individual differences in anthropomorphism questionnaire) (Waytz et al., 2010) to measure people’s ability in anthropomorphism. We can then measure the ERP responses of different groups based on the IDAQ score.

I found neither group differences for either other ERP components in the difference score grouping method or the other two grouping methods, nor systematic patterns of how ERP components differ across groups. Given that the realism effect sustained across different groups on multiple components (predominantly the N170 and P200), it may be possible that individual perception or anthropomorphism is secondary in the processing to the realism of avatars, or at least not involved in early processes. The brain may make an instant judgement of how humanlike these avatars are before a higher-order, conscious perception or discrimination kicks in. Another plausible explanation to the null findings could be the use of the test instrument. In
this analysis, I used a simple survey which was not normed or validated since I did not initially intend to examine participants’ ratings of the faces in any depth. I grouped participants based on their human likeness score from the survey, the grouping therefore may not be systematic and precise enough for showing group effects. Furthermore, the small sample size for each group (13 participants) could reduce our power to detect group effects. Therefore, in future studies, a more validated and standardized questionnaire or survey could be administered to provide a stronger background for grouping participants, and a larger sample size will increase the statistical power to test the group effect. In addition, an alternative approach rather than a quartile split might be to conduct a correlation or regression analysis to maximize the statistical power and avoid issues of misclassification.

5.3.1 Conclusion

This exploratory analysis also provides a direction for further study in investigating if individual differences in anthropomorphizing avatars affect their perceived human likeness of the avatars in terms of ERPs. Although the findings appeared to reject the group differences, one implication of the analysis is to include perceivers’ aspect in designing avatars. We are still unclear whether two individuals with one seeing an avatar more real and another one seeing the same avatar less real, will ultimately differ in their acceptability of the avatar and vulnerability to experience the uncanny valley. However, the possibility of targeting a specific group of perceivers who indicate higher level of acceptability could be an exit to the uncanny valley in avatar or humanoid creation.
Chapter 6 General Discussion

6.1 Main Findings

At the beginning of the thesis, I introduced the Mona Lisa as a metaphor for the ambiguity and enigmas involved in face processing. The ambiguity of her facial expressions and the emotional disengagement viewers experience have inspired people to decipher the myth underlying the masterwork. In a similar manner, the exploratory work in this thesis has focused on whether the AFS produces images that are comparable to photos as stimuli in facial expression studies. I have generated the following findings:

6.1.1 Facial Expression Detection Threshold

I started by investigating the threshold for generating a psychophysiological response to facial expression (Chapter 2). Results suggest that anger and happiness started differentiating at an early stage of visual processing (the N170 time window) at the 20% intensity level. In terms of a single emotion, the intensity required to differentiate from a neutral expression differs between anger and happiness. In particular, happiness required a minimum 40% of the maximum intensity to be distinguished from neutral in the ERP; anger appears to require even higher intensity. The findings require some rethinking or reinterpretation of the three stages of facial expression processing proposed by Luo et al. (2010). Simply put, the three-stage model suggested that threatening emotions (anger and fear) are detected in the first stage, followed by differentiating other emotional expressions from neutral in the second stage, and finally the distinction between emotions occurs in the third stage. In ERPs, these stages progress in temporal order from the N1/P1 through the N170/VPP and finally the N300/P300 time windows. My findings indicated that the entire process of facial expression of emotions may occur early in
processing, as indicated by my observation that the N170 in my study was modulated by both emotion and intensity. I therefore presume that the distinction between emotional and neutral expressions can be initiated as early as the P1 window, in which case between-emotion discrimination occurs approximately at the N170 window.

6.1.2 Hyper-realistic Avatars

If we hope to use hyper-realistic avatars to generate realistic face stimuli with a higher degree of stimulus control that is offered by other types of face stimuli (photos/videos, schematic faces, animations, etc.), it is important to demonstrate that the avatars are processed similarly to photos or videos. If, on the other hand, avatars produce some sort of unique response, it is possible that this response is a manifestation of the uncanny valley. That is, avatar faces may look realistic but nevertheless produce an emotional (or other) response that somehow marks them as different from “real”.

In Chapter 3, I examined whether avatar faces from the AFS work in the same manner as photo faces do. I incorporated a brief survey to generate human-likeness ratings and correct responses of emotion recognition for both avatars and faces. The results suggested that photo and avatar faces did not differ in human-likeness scores, but emotion and intensity have an impact on the perceived humanness of simulated faces. A higher degree of intensity (40% in my study), especially with happy expressions, tends to help push avatars towards the humanness end of a continuum. For correct responses, people were less accurate in identifying emotions on avatar faces, especially 20% angry simulated faces that has the lowest correct response. Overall, these results seem to suggest that the avatars are doing a pretty good job, but there are some clues that they may not be perceived as quite as good/real/human as the photo faces.
The ERP results, nevertheless, exhibited a strong effect for realism differences between photo and simulated faces. The amplitude differences commence around the N170 time window and sustained till 300ms post-stimulus onset. More specifically, both N170 and P200 were larger in amplitude for photo faces than for simulated faces. The findings raise the questions of how prototypical (Leopold et al., 2001) and/or typical (Halit et al., 2000; Schindler et al., 2017) these faces were. Both ideas assume a comparison with a built-in average face represented in the brain, and that any tiny deviation from the average face may trigger a response. If the deviation is enough, the response may manifest as alarm, such as eerie feelings. In addition, simulated faces generated a larger FcEP component, which reflects emotional responses for simulated faces over photo faces. This particular finding may be a manifestation of the emotional element in Mori’s uncanny valley theory.

Furthermore, the study also demonstrated the practicality and reliability of using ERPs to test the realism of hyper-realistic computer-generated avatars in comparison to self-report survey or questionnaires.

6.1.3 Within-avatar Differences

In Chapter 3 I showed evidence that ERPs are sensitive to differences between highly realistic computer-generated avatars and photo faces. The two models I had at the time of experiment, Leah and Xyza, were both created to be as close to human as possible but nevertheless received different feedback on their realism. Leah was generally perceived to be more real than Xyza based on participants’ feedback. In Chapter 4 I was then interested in delving deeper into whether the uncanny valley effect can occur between different avatars. Could one avatar be “uncannier” than the other, and thus show a stronger response? In a wider scope, I also tested
whether ERPs can be a reliable method to detect differences in humanness within avatar models. While the self-report rating of human-likeness score did not seem particularly sensitive to differences between two avatars, the ERP results, revealed a specific component (POz positivity; 120-160ms) that appeared to be sensitive to between-avatar differences. Leah evoked the largest positivity followed by Xyza and photo faces. The N170 did not show differences in amplitude between two avatar models. For the P200, photo faces elicited a magnified amplitude, followed by Leah and Xyza respectively. It is difficult to arrive at a systematic interpretation for these results. If photo faces are used as the baseline of realism, POz positivity would indicate Xyza to be closer to photo, whereas the P200 tends to show a higher degree of realism of Leah. Such dilemma in the interpretation could also be the result of idiosyncratic differences between the two avatars, so my findings are necessarily tentative. Nevertheless, the data suggest that ERP differences in the POz positivity and P2000 windows might be critical for understanding the acceptability of simulated faces.

6.1.4 Individual differences in perceiving avatars

In Chapter 5 I divided the participants based on the human-likeness ratings using different methods (difference score, average score and simulated faces score) and investigated whether individual perception in the degree of realism of the avatars would influence the ERP responses. The findings suggested that those who see simulated faces as more real than photo faces tend to anthropomorphize more and are more accepting of simulations. Although I did not find concrete evidence to further support the group differences, the consideration of perceivers with different anthropomorphism abilities should not be neglected in the creation of avatars.
6.2 Contributions

Based on the work and findings above, this thesis went beyond the current state-of-the-art in the research and understanding of the uncanny valley in several ways.

Empirical work suggested that facial expressions can be distinguished from each other at a relatively low intensity (e.g. 20% level). Furthermore, the degree of intensity required to recognize expressions varies across different emotions. In my study happiness requires 40% intensity and anger requires higher than 40% intensity to be reliably identified. The work provided some evidence for the uniqueness of different emotions and future research could incorporate more emotions (see 6.4 Future Work section for more details).

The second contribution of this thesis is in understanding the mechanism of processing hyper-realistic avatars and the uncanny valley. Despite the advancement in the AI and machine learning to generate avatar faces at an unprecedented level of realism, these stimuli are seldom used as stimuli in research applications. Photo faces have some inherent drawbacks as a stimulus, including posed expressions and oversimplified linear morphs of various intensities. Avatars that are indistinguishable from real human faces would open up an entire new opportunity to exert precise control over face stimuli. Nevertheless, a possible impediment to employing avatars in research could be the lack of empirical studies in investigating the applicability of replacing traditional face stimuli with synthetic faces. Simply put, it remains uncertain whether avatars are as good as photo faces. My studies suggest that they may not be quite there yet. If we want to use such faces to understand emotional expression, it would be better to ascertain that they are not generating unintended emotional responses (uncanny valley). There is enough implication in my data that the processing of avatar faces remains different from that of photo faces to make us
cautious about using avatar faces for more extensive research on emotional expression just yet. Therefore, this thesis, to the best of my knowledge, is the first-of-type in incorporating psychophysiological studies with hyper-realistic avatars (the AFS). The result indicated that despite of visually indistinguishable human-likeness of the AFS from photos, ERP responses illustrated a gap in the realism of the AFS faces. This gap in realism, could contribute to an explanation of the uncanny valley at the psychophysiological level.

Another contribution of the thesis stems from the comprehension in scrutinizing realism by comparing between two avatars and considering from perceivers’ aspect. Further, despite, not finding links between individual perception of avatars’ realism and ERP responses in this research, it still remains a possibility that distinctions in degrees of anthropomorphizing avatars could alter how our brain reacts to avatars. It thus emphasizes the importance of taking perceivers’ aspect into the account while creating the avatars.

All in all, this thesis contributes to the body of work of testing the realism of the AFS and makes some initial steps towards a psychophysiological understanding of the uncanny valley. The work also aids not only in the generation of future hypotheses in exploring avatars as stimuli in emotional and facial perception studies, but also as background research in designing avatars that hopefully do not fall into the uncanny valley.
6.3 Limitation of the studies

6.3.1 Participants

Though the sample sizes for the studies are generally reasonable, the sample size in the first study should preferably be somewhat larger. For the analysis in Chapter 5, as participants from study 2 (Chapter 3) were further divided into groups, the sample size for each group became smaller. Because by subdividing, each group now has small enough numbers that power becomes an issue. Ideally, a sample size of 100 participants would provide the opportunity to subdivide into groups of 25 (100/4) and thus have enough power to draw reliable inferences.

6.3.2 Stimuli and Tasks

Photo faces used in the thesis were derived from the Montreal Set of Facial Displays of Emotion (Beaupre & Hess, 2005). The different intensities of facial expressions were created by morphs that merge 0% and 100% expressions at various combinations. A major issue with the morphs is that in real life facial expressions do not necessarily produce different intensities in a linear manner, nor can emotional intensity be precisely quantified in increments like 20%, or 40%. Moreover, the stimulus faces are monochrome, which do not resemble real life faces.

Apart from that, in this research, I was interested in the processing of lower intensity expressions. Though I found 40% is sufficient in identifying facial expressions, in retrospect adding a higher intensity – say 60% -- could perhaps compensate for brief presentation times and give us a better understanding of the relationships between intensity of expression and strength of behavioral or psychophysiological response. This could be important for the study in Chapter
3 since I failed to find a significant effect of intensity. A more intense expression can be more helpful in understanding any interaction between realism and intensity of facial expression.

Another limitation is the use of static emotional faces as stimuli. In reality, facial expressions are dynamic. Although the stimuli displayed a range of intensities, they were static images, and lacked the sense of movement that animates real faces. Earlier studies (Ambadar, Schooler, & Cohn, 2005; Bould & Morris, 2008) have revealed that the recognition of dynamic facial expressions outperforms that of static expressions, especially for subtle expressions. In addition, motion also plays an important role in the experience of the uncanny valley (McDonnell et al., 2012; Piwek et al., 2014). Of course, most instances of virtual agents applied in real life are also dynamic.

The survey used in this thesis needs to be further expanded, validated and normed. This is important since the grouping methods in Chapter 5 are based on the survey score. A more valid and comprehensive survey might allow grouping to be more accurate, and thus may reveal stronger group effects.

The most important limitation in this research lies in the very limited number of avatars that were available for our use. Both avatars are female, and they differ in apparent age, attractiveness, and along any number of other dimensions. This may hamper the generalization of the findings, and is a particular issue for avatar models’ comparison in Chapter 4. In the future, once more avatars have been created, it will be important to extend this work with as wide a range of avatars as possible.
6.4 Future work

The findings in this thesis provide a large scope for future research to develop hyper-realistic synthetic faces.

The use of hyper-realistic, computer–generated avatars as tools for psychophysiological research offers nearly endless opportunities. Once a larger number of such avatars becomes available, I expect to be able to offer a highly controllable, hyper-realistic testing environment that will allow far more nuanced questions than current stimuli provide. In order to achieve this goal, however, we must be able to understand, predict, and control the aspects of the stimuli that result in unintended uncanny-valley responses. Once such control is achieved, we may be able to create faces with variations in attractiveness or age without necessarily compromising other dimensions such as identity. With avatars, we can control details of individual muscle groups to generate facial expressions that resemble – or depart from – those seen in real life.

Another line of future work is to advance the research into the uncanny valley. The concept was developed in 1970s, but it was not studied comprehensively until the recent emergence of AI. It is an ultimate goal for many people including computer scientists, engineers, psychologists and businessmen to create a virtual agent with unprecedented realism that is not emotionally rejected by users. Therefore, a direction of future work can explore the psychophysiological and neurological signals of the uncanny valley experience using a wide range of techniques, including but not limited to EEG and fMRI.

In the broader context of facial expressions processing, future studies should explore the detection thresholds of more emotions to investigate how the brain requires different intensity
levels to identify different emotions from neutral. It is also important to extend the findings of the thesis to investigate the applicability of emotional quantification that could systematize and generalize emotion studies using different intensities. For instance, instead of using 20% or 40% intensity expressions that may differ across studies for subtle expressions, we can define differences in expressions by the just noticeable difference (JND) required to produce modulations of the N170 (or other ERP components) responses as subtle expressions. This, of course, requires further studies to establish a more consolidated system.

6.5 Conclusion

An important motivation of the current thesis was to test of the AFS is able to, or anywhere close to pass the facial version of Turing Test. While being in principle a preliminary start of psychophysiological research of hyper-realistic avatars, the work in the thesis has shaped, and continues to shape the practicality of incorporating synthetic faces in psychological research and the understanding of the uncanny valley to move forward in passing the facial Turing test.
References


doi:https://doi.org/10.1080/1354680044400245


Interactive Communication (RO-MAN), (pp. 368-373). Jeju, Korea.
doi:10.1109/ROMAN.2007.4415111


doi:https://doi.org/10.1371/journal.pone.0026551


220


224


doi:https://doi.org/10.1371/journal.pone.0003628


doi:https://doi.org/10.1016/j.neuropsychologia.2006.04.022


doi:https://doi.org/10.1016/j.cognition.2012.06.007


doi:https://doi.org/10.1016/j.neuropsychologia.2017.10.016


doi:10.1002/cne.23368


doi:10.4018/jse.2010101605


doi:10.1080/17470919.2015.1059361


251


https://springfielddaily.com/technology/nvidia-creates-an-ai-that-designs-exact-faces-to-the-real-ones/

Squires, N. (2015, August 20). *Have researchers unlocked the mystery of Mona Lisa’s famously enigmatic smile?* Retrieved 2 2, 2018, from The Telegraph:


doi:https://doi.org/10.1016/j.visres.2011.08.024

258


259


doi:https://doi.org/10.1016/j.neuropsychologia.2012.02.030
Appendix A

Humanness Survey

1. On a scale from 1 (Not at all) to 10 (Very much), do you think this is a photo of a real person?

2. On a scale from 1 (Not at all) to 10 (Very much), please rate the “humanness” of the picture.

3. On a scale from 1 (Not at all) to 10 (Very much), does the picture look weird to you?

4. On a scale from 1 (Not at all) to 10 (Very much), does the emotion depicted by the face look “real” to you?

5. On a scale from 1 (Not at all) to 10 (Very much), are you able to recognize the emotional expressions on the face? What is the expression?

6. On a scale from 1 (Not at all) to 10 (Very much), do you think the person shown in the image is capable of understanding others’ feeling?