THE UNIVERSITY OF

SHEFFIELD

CONTEXTUAL RECOGNITION OF ROBOT EMOTIONS

JIAMING ZHANG

March 2013

Submitted for the Degree of Doctor of Philosophy

Department of Computer Science

The University of Sheffield

Co-Supervisors: Dr. Amanda Sharkey

Prof. Noel Sharkey

Declaration

I declare that this report is composed by myself and the work contained herein is my own except where explicitly stated otherwise in the text. This work has not been submitted for any other degree or professional qualification except as specified.

Jiaming Zhang

Dedicated to my deceased family members

Grandmother Mrs. Ruifang Luo and

Grandfather Mr. Shanzhong Peng and

My deceased high school classmate Mr. Zhoulong Zhang

Abstract

In the field of human-robot interaction, socially interactive robots are often equipped with the ability to detect the affective states of users, the ability to express emotions through the use of synthetic facial expressions, speech and textual content, and the ability for imitating and social learning. Past work on creating robots that can make convincing emotional expressions has concentrated on the quality of those expressions, and on assessing people's ability to recognize them. Previous recognition studies presented the facial expressions of the robots in neutral contexts, without any strong emotional valence (e.g., emotionally valenced music or video). It is therefore worth empirically exploring whether observers' judgments of the facial cues of a robot would be affected by a surrounding emotional context. This thesis takes its inspiration from the contextual effects found on the interpretation of the expressions on human faces and computer avatars, and looks at the extent to which they also apply to the interpretation of the facial expressions of a mechanical robot head. The kinds of contexts that affect the recognition of robot emotional expressions, the circumstances under which such contextual effects occur, and the relationship between emotions and the surrounding situation, are observed and analyzed in a series of 11 experiments. In these experiments, the FACS (Facial Action Coding System) (Ekman and Friesen, 2002) was applied to set up the parameters of the servos to make the robot head produce sequences of facial expressions. Four different emotional surrounding or preceding contexts were used (i.e., recorded BBC News pieces, selected affective pictures, classical music pieces and film clips). This thesis provides evidence that observers' judgments about the facial expressions of a robot can be affected by a surrounding emotional context. From a psychological perspective, the contextual effects found on the robotic facial expressions based on the FACS, indirectly support the claims that human emotions are both

biologically based and socially constructed. From a robotics perspective, it is argued that the results obtained from the analyses will be useful for guiding researchers to enhance the expressive skills of emotional robots in a surrounding emotional context. This thesis also analyzes the possible factors contributing to the contextual effects found in the original 11 experiments. Some future work, including four new experiments (a preliminary experiment designed to identify appropriate contextual materials and three further experiments in which factors likely to affect a context effect are controlled one by one) is also proposed in this thesis.

Acknowledgements

For the PhD thesis I presented here, first and foremost, I would like to express my sincere gratitude to my co-supervisors: Dr. Amanda Sharkey and Prof. Noel Sharkey, especially Dr. Amanda Sharkey, without whom I would never receive so much enthusiastic advice, support and guidance at every stage during the PhD period. All the participants in my experiments, without whom no sufficient data could be collected and this thesis would not be complete.

I am grateful to the Department of Computer Science in the University of Sheffield. Here I was trained to be a researcher and was offered the job of a teaching assistant. It has provided many opportunities for my professional development. I would like to thank a few people whom I am very much indebted to. Dr. Gordon Manson, who helped me to replace the Tinibox with Java Sun Spot to enable the mechanical robot head CIM to be emotional, has provided me crucial technical support. Dr. Sifat Momen and Dr. Sanaz Jabbari, former PhD students in the same lab, have given me lots of helpful advice for my study and life in Sheffield.

Furthermore, I would like to thank my friends. In particular, Mr. Weiwei Chen, Mr. Junyu Wang and Dr. Jionglong Su have inspired me with creative ideas through our many conversations. I am also indebted to all other friends who have offered me support without which my life in the UK would be arduous. Dr. Yi Li and his adorable family, who keep inviting me to their home regularly and treating me hospitably, have substantially enriched my life here. Last but not least, my lovely family in China, from which I receive generous financial and spiritual support at all times.

Table of Contents

CHAPTER 1: INTRODUCTION 16

1.1 Robotic Facial Expressions: from Psychology to Technology 17

1.2 Emotion and Context 19

1.3 Thesis Objectives 23

1.4 Contributions and Relevant Publications 25

1.5 Organization of the Thesis 27

CHAPTER 2: PSYCHOLOGY OF EMOTION 29

2.1 Classical Theories 29
2.1.1 Evolutionary Theories 29
2.1.2 Cognitive-appraisal Theories 36
A. Appraisal Theories of Emotion 36
B. Comparing Evolutionary and Appraisal Accounts 41
2.1.3 Social Constructionist Theories 43

2.2 Novel Approaches 47
2.2.1 Theory of Embodying Emotion 47
2.2.2 Embodied Appraisal Theory 50
2.2.3 Social Functionalist Approach 52
2.2.4 Perceptual Control Theory on Emotions 54
2.2.5 Rolls's Theory of Emotion 56

2.3 Summary and Conclusion 61

CHAPTER 3: SYNTHETIC ROBOT FACIAL EXPRESSIONS 65

3.1 Social Robot Embodiment 67
3.1.1 Humanoids 68
3.1.2 Androids 68
3.1.3 Creature-like Social Robots 69
3.1.4 Non-humanoid and Non-zoomorphic Social Robots 70

3.2 Creating Recognizable Facial Expressions for Social Robots 71
3.2.1 Examples of Non-Android Robots 73
A. Kismet 73
B. Probo 77
3.2.2 Examples of Android Robots 81
A. Geminoid F 81
B. Jules 84

3.2.3 Summary of the Methods 86

3.3 "Social" Aspects of the Design of a Human-robot Interaction 88
3.3.1 Robot Attributes 88
3.3.2 User Attributes 90
3.3.3 Task Structure 91

3.4 Summary and Conclusion 91

CHAPTER 4: CONTEXTUAL EFFECTS ON JUDGMENTS ABOUT SYNTHETIC ROBOT FACIAL EXPRESSIONS 96

4.1 Generic Experimental Method 96
4.1.1 The Motivation of This Thesis 96
4.1.2 The Hypotheses 99
4.1.3 The Dependent Variable 100
4.1.4 The Independent Variable 102
4.1.5 The Experimental Conditions 103
4.1.6 The Number of the Experiments 105
4.1.7 The Generic Experimental Procedures 107
4.1.8 The Statistical Analysis Techniques 109

4.2 Experiment Speech1 110
4.2.1 Experimental Results 112
4.2.2 Discussion and Conclusion 115

4.3 Experiment Speech2 120
4.3.1 Experimental Results 121
4.3.2 Discussion and Conclusion 124

4.4 Experiment Image1 125
4.4.1 Experimental Results 127
4.4.2 Discussion and Conclusion 129

4.5 Experiment Image2 130
4.5.1 Experimental Results 131
4.5.2 Discussion and Conclusion 134

4.6 Experiment Image3 135
4.6.1 Experimental Results 135
4.6.2 Discussion and Conclusion 138

4.7 Experiment Music1 138
4.7.1 Experimental Results 140
4.7.2 Discussion and Conclusion 143

4.8 Experiment Music2 144
4.8.1 Experimental Results 145
4.8.2 Discussion and Conclusion 149

4.9 Experiment Music3 150
4.9.1 Experimental Results 151
4.9.2 Discussion and Conclusion 154

4.10 Experiment Video1 **155**
4.10.1 Experimental Results 156
4.10.2 Discussion and Conclusion 158

4.11 Experiment Video2 **159**
4.11.1 Experimental Results 160
4.11.2 Discussion and Conclusion 163

4.12 Experiment Video3 **164**
4.12.1 Experimental Results 165
4.12.2 Discussion and Conclusion 167

4.13 Summary and Conclusion **167**

CHAPTER 5: AN ANALYSIS OF THE METHODOLOGY **180**

5.1 A Further Experiment **180**
5.1.1 The Motivation of the Experiment 180
5.1.2 The Experimental Conditions 181
5.1.3 Experimental Procedures 181
5.1.4 Experimental Results 183
5.1.5 Conclusion and Discussion 189

5.2 Possible Factors Contributing to the Contextual Effects Found in the Previous 11 Experiments **190**
5.2.1 The Choices of the Contexts 190
A. The Content of the Contexts 191
B. The Source Clarity of the Contexts 192
C. The Difference Between One Modality and Multiple Modalities 195
5.2.2 The Rating Scheme 195
5.2.3 The Mood Effect 197
5.2.4 The Manner of Presenting the Robot Face and the Context 197
A. The Style of the Presentation of a Robot Face and a Context 198
B. The Contrast Effect 199

5.3 Summary and Conclusion **200**

CHAPTER 6: PROPOSED NEW EXPERIMENTS ADDRESSING THE IMPACT OF FILM CLIPS **202**

6.1. The Preliminary Experiment **203**
6.1.1 The Motivation of the Preliminary Experiment 203
6.1.2 The Experimental Conditions 206
6.1.3 The Generic Experimental Procedures 206
6.1.4 The Statistical Analysis Techniques 208
6.1.5 Predictions of the Experimental Results 208

6.2. The Proposed Three New Experiments **211**
6.2.1 The Motivation of the New Experiments 211
6.2.2 The Hypotheses 214
6.2.3 The Dependent Variable and the Independent Variable 215
6.2.4 The Experimental Conditions 216
6.2.5 The Generic Experimental Procedures 217

6.2.6 The Statistical Analysis Techniques 218
6.2.7 Predictions of the Experimental Results 219

6.3 Summary 221

CHAPTER 7: SUMMARY AND FUTURE WORK 223

7.1 Summary of the Thesis 224

7.2 Future Work 228

REFERENCES 231

APPENDIX Ⅰ: A LIST OF THE ACTUAL ORDER OF THE 11 EXPERIMENTS 249

APPENDIX Ⅱ: DETAILED EXPERIMENTAL SET UPS FOR ALL THE 11 EXPERIMENTS 250

APPENDIX Ⅲ: THE QUESTIONNAIRE OF THE EXPERIMENT SPEECH1 251

APPENDIX Ⅳ: THE PROCEDURES OF EXPERIMENTS IMAGE1 AND MUSIC3 CONDUCTED IN PAIRS 260

APPENDIX Ⅴ: THE PROCEDURES OF EXPERIMENTS IMAGE3 AND VIDEO2 CONDUCTED IN PAIRS 262

APPENDIX Ⅵ: THE PROCEDURES OF EXPERIMENTS VIDEO1 AND MUSIC2 CONDUCTED IN PAIRS 264

APPENDIX Ⅶ: THE PROCEDURES OF EXPERIMENTS MUSIC1 AND IMAGE2 CONDUCTED IN PAIRS 266

APPENDIX Ⅷ: THE PROCEDURES OF THE EXPERIMENTS VIDEO3 AND SPEECH2 CONDUCTED IN PAIRS 268

APPENDIX Ⅸ: THE PROCEDURES OF THE PRELIMINARY EXPERIMENT 270

APPENDIX Ⅹ: THE QUESTIONNAIRE OF EACH PROPOSED NEW EXPERIMENT 272

List of Figures

Fig. 2.1: Comprehensive illustration of the CPM of emotion (Sander et al. 2005), permission to reprint granted. 40

Fig. 2.2: Some of the emotions associated with different reinforcement contingencies are indicated. Intensity increases away from the center of the diagram, on a continuous scale. The classification scheme created by the different reinforcement contingencies consists with respect to the action of (1) the delivery presentation of a reward (S+), (2) the presentation of a punisher (S–), (3) the omission of a reward (S+) (extinction) or the termination of a reward (S+!) (time out), and (4) the omission of a punisher (S–) (avoidance) or the termination of a punisher (S–!) (escape) (from (Rolls, 2007), permission to reprint granted. 58

Fig. 2.3: The connections between the three classical theories and other novel approaches to emotion (the arrow indicates the direction of inspiration, for instance, Rolls's Theory of Emotion combines insights both from Evolutionary Theories and Cognitive-appraisal Theories). 62

Fig. 3.1: This diagram illustrates where the basis postures are located in affect space. Permission from Emotion and Sociable Humanoid Robots (Breazeal, 2003) to reprint granted. 74

Fig. 3.2: The 6 basic facial expressions of the covered prototype used in T7 (happy, surprise, sad, anger, fear and disgust). Permission from (Saldien et al. 2010) to reprint granted. 79

Fig. 3.3: The android robot Geminoid F (top row) and its model person (bottom row) portraying five basic emotions and a neutral expression, from left to right: angry, fearful, happy, neutral, sad, surprised. Copyright © 2011 IEEE, reprinted with permission from (Becker-Asano and Ishiguro, 2011). 83

Fig. 3.4: Well mapped mouth shapes: The top and bottom of each row show theactress's facial expression and the resulting poses of the robot head, respectively. Reprint with permission from (Jaeckel et al. 2008b). 85

Fig. 4.1: Joy (top left), Surprise (top middle), Fear (top right), Sadness (bottom left), Anger (bottom middle), and Disgust (bottom right) of CIM. Note that the robot head neither have lid nor have jaw resulting in no AU 7 or AU 26, for instance, in use. 102

Fig. 4.2: Grouping of subjects with a surrounding context 104

Fig. 4.3: Grouping of subjects with a preceding context 104

Fig. 6.1: The Geneva Emotion Wheel (Scherer, 2005) which has 20 emotion items (families) which themselves have double labels. Five scales are available for the intensity rating of each item, and "No emotion" and "Other emotion" fields increases the self-report options. 213

Fig. 6.2: The same grouping of each new experiment 217

Fig. Appendix Ⅵ: Experimental Instructions of the two experiments Video1 and Music2 264

Fig. Appendix Ⅶ: Experimental Instructions of the two experiments Music1 and Image2 266

Fig. Appendix Ⅷ: Experimental Instructions of the two experiments Video1 and Speech2 268

List of Tables

Table 2.1 Relationships between the functions and components of emotion, and the organismic subsystems that subserve them (Sander et al. 2005), permission to reprint granted. 40

Table 4.1 Servos set up for CIM. 102

Table 4.2 The 11 Experiments presented in this thesis 105

Table 4.3 Content of the recorded BBC World News. 111

Table 4.4 Percentage of six possible labels (items in the first row) chosen to match the displayed facial expressions (items in the first column) in E1. 112

Table 4.5 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Speech1. 113

Table 4.6 Percentage of three possible labels (items in the first row) chosen to match the played sequence of BBC News items (positive/neutral or negative) in Speech1. 114

Table 4.7 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Speech2. 122

Table 4.8 Percentage of three possible labels (items in the first row) chosen to match the played sequence of BBC News items (positive/neutral or negative) in Speech2. 123

Table 4.9 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Image1. 127

Table 4.10 Percentage of three possible labels (items in the first row) chosen to match the displayed set of affective pictures (positive or negative) in Image1. 128

Table 4.11 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Image2. 132

Table 4.12 Percentage of three possible labels (items in the first row) chosen to match the displayed set of affective pictures (positive or negative) in Image2. 133

Table 4.13 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Image3. 137

Table 4.14 Percentage of three possible labels (items in the first row) chosen to match the displayed set of affective pictures (positive or negative) in Image3. 137

Table 4.15 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Music1. 141

Table 4.16 Percentage of three possible labels (items in the first row) chosen to match the played piece of classical music (happy or sad) in Music1. 142

Table 4.17 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Music2. 146

Table 4.18 Percentage of three possible labels (items in the first row) chosen to match the played piece of classical music (happy or sad) in Music2. 148

Table 4.19 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Music3. 152

Table 4.20 Percentage of three possible labels (items in the first row) chosen to match the played piece of classical music (happy or sad) in Music3. 153

Table 4.21 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Video1. 157

Table 4.22 Percentage of three possible labels (items in the first row) chosen to match the displayed film clip (amused (happy) or sad) in Video1. 158

Table 4.23 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Video2. 161

Table 4.24 Percentage of three possible labels (items in the first row) chosen to match the displayed film clip (amused (happy) or sad) in Video2. 162

Table 4.25 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Video3. 166

Table 4.26 Percentage of three possible labels (items in the first row) chosen to match the displayed film clip (amused (happy) or sad) in Video3. 166

Table 4.27 Findings of the 11 experiments including the gender difference and cross-cultural difference (examined by Fisher's Exact Test in PASW 18) 168

Table 4.28 Recognition accuracy (%correct) of each emotional context in the 11 experiments 175

Table 5.1 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative). 186

Table 5.2 Percentage of three possible labels (items in the first row) chosen to match four other emotional materials 187

Table 6.1 Choices of film clips and musical pieces provided in the preliminary experiment 205

Table 6.2 Expected results of the preliminary experiment 210

Table Appendix Ⅰ The actual order of the 11 experiments 249

Table Appendix Ⅱ More experimental set ups for all the 11 experiments 250

Chapter 1: Introduction

Suppose you were on your way to a museum that had a robot receptionist. On your journey, you listen to a very sad piece of classical music. When you enter the museum, the robot receptionist greets you with a tentative smile, just as a news broadcast comes over the speakers announcing that the football club you hate has just won the championship. When you are asked what the robot receptionist is feeling you answer “sad”. Why? The music and football news have influenced your view of the robot’s expressions. In the future, it is quite likely that human beings will encounter some real-life situations where a robot that produces emotional expressions will be surrounded by unexpected contradictory emotional contexts.

Robots are beginning to be used in social situations where there is a need for them to seem to respond appropriately to humans. It will be easier to accept robot receptionists and robot guides that seem to respond to us with appropriately expressed facial emotions. There is a growing interest in human-robot interaction, but its researchers have tended to focus on the capabilities of sociable emotional robots themselves, for example, their ability to detect the affective states of users, their ability to express emotions through the use of synthetic facial expressions, speech and textual content, and their ability for imitating and social learning.

Since facial expression is a major modality in human face-to-face communication (Mehrabian, 1968), researchers in the field of human-robot interaction focus a lot on how to create convincing and believable synthetic facial expressions of socially interactive robots. Generally speaking, socially interactive robots share a unifying characteristic that they engage people in an interpersonal manner, communicating and coordinating their behavior with humans through verbal, nonverbal, or affective modalities (Breazeal et al. 2008).

Although robots can vary from humanoid, android, creature-like, to simply non-humanoid and non-zoomorphic in form, it has been possible to create convincing and believable synthetic emotional expressions for different kinds of robots, for instances, Ishiguro' new android robot "Geminoid F" (Becker-Asano and Ishiguro, 2011) and Bristol Robotics Laboratory's BERT2 (Bazo et al. 2010) which has a hybrid face with affective facial expression implementations.

1.1 Robotic Facial Expressions: from Psychology to Technology

There are different theories of emotion focusing on different aspects of, or questions about emotion:

(1) What is an emotion? For example, Evolutionary Theories (e.g., Darwin, 1872/1998; Tomkins, 1962; 1963; Ekman et al. 1969; Ekman, 1999; Cosmide and Tooby, 1987; 2000) and Social Constructionist Theories (e.g., James, 1890; Schachter and Singer, 1962; Russell, 2003; Barrett, 2006) mainly focus on this aspect.

(2) How is an emotion generated? For instance, Cognitive-appraisal Theories (Arnold, 1960; Lazarus, 1966, 1991; Scherer 1984, 2001; Smith and Ellsworth 1985; Roseman and Smith, 2001; Sander et al. 2005), Theory of Embodying Emotion (e.g., James, 1884; Lange, 1885; Damasio, 1989, 1994; Barsalou, 2003; Barsalou et al. 2003; Chartrand and Bargh, 1999; Niedenthal et al. 2005; Vermeulen et al. 2007; Niedenthal et al. 2007; Niedenthal, 2009;), Embodied Appraisal Theory (e.g., James, 1884; Lange, 1885; Lazarus, 1991; Damasio, 1999; Prinz, 2004a, 2004b), Perceptual Control Theory on Emotions (e.g., Dewey, 1896; Powers et al. 1960; Powers, 1973, 2007, 2008; Moore, 2007), and Rolls's Theory of Emotion (Millenson, 1967; Weiskrantz, 1968; Gray, 1975, 1987; Rolls, 1986a, 1986b, 1990, 1999a, 2000a, 2005a, 2007) mainly focus on this aspect.

(3) Why do humans and other species have emotions? For example, Social Functionalist Approach (Keltner et al. 2006) mainly focuses on this aspect, Evolutionary Theories, Social Constructionist Theories and Rolls's Theory of Emotion only partially focus on this aspect.

According to Breazeal et al. (2008), in order to be able to participate effectively in emotion based interactions, robots must be equipped with abilities to recognize and interpret affective signals from humans, possess their own internal models of emotions (often inspired by psychological theories), and be able to communicate this affective state to others. These internal models of emotion are of critical importance when interacting socially with humans (Breazeal, 2003a), as the robot's computational model of emotion determines the robot's emotional responses according to its interactions with the external environment and its own internal cognitive-affective state (Breazeal et al. 2008). Therefore, after socially interactive robots are given rich facial features, the insights from the psychology of emotion and its expression are often taken to design their computational models of emotion. Actually, seemingly conflicting theories of emotion have been adapted to design computational models of emotion for some emotional robots. For instance, the FACS (Facial Action Coding System) (Ekman and Friesen, 1978; 2002) which is a product of Evolutionary Theories, and Russell's circumplex model of affect containing two dimensions: valence and arousal (Russell, 1997; Posner et al. 2005), which is a kernel of Social Functionalist Approach, were adapted and combined successfully to design computational models of emotion for MIT's Kismet (Breazeal, 2002) and Vrije Universiteit Brussel's Probo (Saldien et al. 2010). The FACS was found to be a reliable tool for creating at least six basic and discrete facial expressions (joy, sadness, anger, fear, disgust and surprise) for Kismet and Probo, and Russell's circumplex model of affect was adapted to map their discrete facial expressions into an emotional space allowing dramatic but smooth shifts from one emotional state of a robot to another.

Kismet and Probo are successful examples of the applications of insights from the psychology of emotion. These insights, which gave birth to different computational models of emotion, serve important roles in human-robot interaction. At the same time, alternative techniques such as using certain mathematical models which map from facial expressions of human beings to a robot's muscle or servo space are worth probing. For instance, with the advantage of a very human-like appearance, some androids such as Bristol Robotics Laboratory's Jules (Jaeckel et al. 2008a; 2008b; Henrik et al. 2009), were equipped with good facial expressive skills by simply mapping from video footage to the robot head. The Jules project demonstrated that, good quality of mapping that allows the robot to automatically mimic a person's expressions simply by watching their face through a video camera, could be achieved.

1.2 Emotion and Context

Previous work on creating robots that can make convincing emotional expressions has concentrated on the quality of those expressions, and on assessing people's ability to recognize them. For instance, reported results showed that people are able to recognize the emotional expressions of Kismet (Breazeal, 2002) and Probo (Saldien et al. 2010). Such evaluations are useful, and show that it is possible to develop robots that can form recognizable emotional expressions.

It seems that people can attribute the basic emotions to the facial expressions of some emotional robots to an extent. This could imply that the best approach to creating convincing robot expressions would be to assume the universality of synthetic robot facial expressions by using the FACS, and to increase their effectiveness by adapting Russell's circumplex model. However, cross-cultural differences have been found in attributions of specific emotions to facial expressions (Russell, 1991; Russell, 1994; Haidt and Keltner, 1999; Elfenbein and Ambady, 2002). Furthermore, contexts can also play a significant role in people's perception of human emotions. Cowie (2009) concluded that there are two types of context effect: one is that context can affect classifications of emotion subtler than the prototypical categories, the other one is that context can affect people's identification of the object of an emotion. He also gave an example of a second type of context effect, as found in Cauldwell (2000)'s study, that knowing the speaker's habitual settings resulted in listeners' identification of an angry voice with a context as a neutral voice, whilst the same speech sample of anger in isolation evoked reliable impressions.

There is a growing body of empirical research that has begun to demonstrate that the surrounding context can influence the perception and recognition of the facial emotional expressions of human beings. For example, in a particular section called "What Information Determines the Recognition of Emotion?" of the book "The Psychology of Emotion: Interpersonal Experiential, and Cognitive Approaches", Niedenthal et al. (2006b) concluded that there are many factors coloring the recognition of emotions, such as the emotion-eliciting situation in which the expression occurs, the observer's own emotional state, and the types of facial expression that the observer has encountered previously that may result in contrast effects. To begin with, they concluded that both the human face and the surrounding context make important contributions to emotion judgments. In more detail, the way in which facial and context cues are combined (congruent or incongruent with each other), affected observers' judgments of facial expressions. They explained that when the surrounding emotional context did not match the facial expression, observers often reinterpreted either the facial expression (i.e., the person's face does not reveal the person's real feelings) or the meaning of the context (i.e., the situation does not have the usual meaning for the person). Their conclusions were based on studies by Watson (1972), by Wallbott (1988), and by Carrera-Levillain and Fernández-Dols (1994), which showed that contextual effects on the recognition of human emotional expressions are stronger when the expressions are ambiguous or natural and less intense. In a second set of studies by Schiffbauer (1974), and by Leppanen and Hietanen (2003), it was found that observers of human facial expressions were influenced by their own emotional state in their attributions of emotional states to the people they observed. Last but not least, there is evidence that previously encountered facial stimuli can color the recognition of facial expressions of human beings. There is evidence of contrast effects in which previously shown contrasting facial stimuli were found to not only increase the salience of the target human expression (Thayer 1980a; 1980b) but also to qualitatively change the attribution of the emotion category label especially if the judged expression is somewhat ambiguous (Russell and Fehr, 1987). A case in point of Russell and Fehr's findings was that observers who previously saw a low arousal anchor expression (i.e., sadness) actually perceived an anger target expression as anger, while they perceived the same target expression as sad when a high arousal expression (i.e. fear, anger or disgust) preceded it.

Izard (1971) and Ekman (1972) proposed a facial dominance account of human emotions that assumed that a clear prototypical facial expression of a basic emotion would override any expectation derived from the surrounding situation. However, an increasing amount of evidence supports an account of limited situational dominance, whereby facial expressions are seen as only one of the elements contributing to the emotion attributed to a person, and the surrounding situation is found to have a strong effect. For example, Carroll and Russell (1996) reported a study in which they found that even when facial expressions were of a basic emotion, the context in which they occurred determined the specific emotion believed to be expressed by the expresser. In their study, the surrounding situation affected the interpretation of the facial expression, but sometimes the effects are bidirectional. For instance, when De Gelder and Vroomen (2000) explored the combination of information from human facial expressions and voice tone, they found bidirectional contextual effects. They presented a series of pictures of faces ranging from happy to sad, accompanied by a sad voice, a happy voice or no voice. When the emotions matched, people's reaction time in judgments was faster than when they were mismatched. The vocally expressed emotions affected judgments of the facial emotions, and vice versa.

Context has also been found to affect people's interpretation of the emotional expressions of computer avatars (Creed and Beale, 2008; Noël et al. 2009;

Mower et al. 2008; 2009). These studies based the avatar’s facial expressions on the FACS, with the exception of Creed and Beale who based their facial expressions on other work by Ekman (2004). Creed and Beale (2008) investigated how mismatched facial and audio expressions were perceived, combining an animated avatar face with a female human voice. They found that subjects attempted to make mismatched expressions consistent on both the visual and audio dimensions of animations with the result that their perception of the emotional expressions became confused. Noël et al. (2009), on the other hand, found that recognition of the emotional expressions of an avatar was not improved by the presentation of congruent, rather than incongruent emotional text. However the texts they used were very short, and their emotional expressions were preselected on the basis of being extremely recognizable.

In a study similar to Creed and Beale’s, Mower et al. (2008; 2009) used computer simulated avatars in an effort to determine how participants made emotional decisions when presented with both conflicting (e.g. angry avatar face, happy avatar voice) and congruent information (e.g. happy avatar face, happy avatar voice) in an animated display consisting of two channels: the facial expression and the vocal expression. They found that: (1) when presented with a congruent combination of audio and visual information, users were better able to differentiate between happy and angry emotional expressions than when they were presented with either of the two channels individually; (2) when faced with a conflicting emotional presentation, users predominantly attended to the vocal channel rather than the visual channel. In other words, they were more likely to see the avatar’s expressed emotions as being expressed by means of its voice, than reflected in its facial expressions. Their findings indicate that the observer’s recognition of facial expressions can be influenced by the surrounding context, and also that emotion conveyed by other modalities, in this case the voice, can override that expressed by the face.

Empirical research on human and avatar emotions shows that people’s attribution of specific emotions to facial expressions is susceptible to situational information. This research goes against the idea of universally recognisable human facial expressions. The recognition studies using Kismet and Probo, presented the facial expressions of the robots without any non-neutral surrounding contexts with overriding emotional valence (e.g., valenced music or video). Will situational information similarly influence the recognition of synthetic robot facial expressions, even when they are based on the FACS system? Since synthetic robot facial expressions are more ambiguous than human facial expressions, it seems likely that the recognition of the simulated emotions shown by a robot will be similarly influenced by a surrounding affective context.

1.3 Thesis Objectives

As I did not focus on determining the robot's emotional responses according to its interactions with the external environment and its own internal cognitive-affective state in my experiments, there was no need to design a computational model of emotion such as those used in the Kismet and Probo projects. Also, as the robot used in my experiments lacks a human-like appearance, I did not embrace any mapping technique such as that used in the Jules' project. What I tried to investigate was the influence of an emotional context on the emotional judgments of the facial expressions of a robot. As a first step of the exploration, I created readable facial expressions of a robot face with the help of FACS (Facial Action Coding System) (Ekman and Friesen, 2002). Then I looked into the effects of each of the emotional contexts, namely the News recordings, selective affective pictures, the musical pieces, and the film clips, on the recognition of sequences of robotic facial expressions.

Therefore, this thesis looks into the field of emotional judgments of facial expressions of a robot head in a given emotional surrounding context. Similar to the Kismet and Probo project, the FACS was applied to set up the parameters of the servos to make a mechanical robot head produce a long sequence of some or all of the six distinct facial expressions (joy, fear, surprise, anger, disgust, and sadness). The robot head is quite mechanical-like, lacking skin and hair and having limited degrees of freedom (see Section 4.1.3 in Chapter 4 for more details). The moving robot head displayed a sequence of emotional expressions that could be described as either positive, or negative, and did so at the same time as an accompanying surrounding context (e.g., congruent or incongruent recorded BBC News, or selected affective pictures, or classical music or film clips). A robot's facial expressions can be viewed as a modality containing emotional information. At the same time, a surrounding or preceding context, such as a piece of classical music that is either happy or sad, can be treated as another modality. The two modalities may reflect congruent or incongruent emotions. In the present thesis, seven experiments were designed to discover whether different preceding or accompanying contexts would influence people's interpretation of a robot's emotional expressions (see Chapter 4 and 5 for more details).

The objectives of this thesis are fourfold: (1) to see whether the recognition of robot emotional expressions can be affected by context; (2) to see what kinds of context are likely to affect the recognition of robot emotional expressions; (3) to see the circumstances under which such contextual effects are likely to occur; (4) and to see whether when such effects occur, they are always dominant or can sometimes be bidirectional such that the robot's emotional expressions can reversely affect the recognition of the surrounding contexts. I realize the first objective by combining the facial cues of a robot and the contextual cues (congruent or incongruent with each other) to see if the contextual cues could affect subjects' judgements about the synthetic robot facial expressions. The second objective of the thesis is approached by analyzing the the difference between different types of surrounding contexts to see if every kind of emotional contexts always produces a contextual effect. The third objective of the thesis is approached by investigating whether the manner in which the context and the robot head are presented (simultaneous or separate), could affect observers' judgements about synthetic robot facial expressions. In some situations, observers are presented with conflicting emotional cues from both the robot face and the context. To realize the fourth objective, the relative weight of the surrounding context and the robot face on the emotional judgments of the robot face is examined in these situations. Meanwhile, it is also interesting to see whether the observers' internal context, e.g., their mood states, and whether observers have seen an opposite valence of robot facial expressions before the presentation of the robot face (i.e., on second viewing of the robot face), would color their attribution of the facial expressions of the robot to specific emotions.

1.4 Contributions and Relevant Publications

The following list highlights the main contributions of the thesis:

(1) This thesis provides evidence that the emotional judgments of the facial expressions of a robot can be affected by a given emotional surrounding context. It contributes to the psychology of emotion, which has already observed that people's interpretation of human emotional expressions and of computer avatars can be affected by the surrounding emotional context. In more detail, this thesis shows that, first of all, not all the surrounding contexts have contextual effects; secondly, the manner in which the context and the robot head are presented (simultaneous or separate) is important in obtaining such a contextual effect; thirdly, the way in which the facial cues of a robot and the contextual cues were combined (congruent or incongruent with each other), can affect subjects' judgements about synthetic robot facial expressions; fourthly, a situational dominance is not found for every surrounding context having a contextual effect; finally, the issue of the relationship between emotions and the surrounding situation is a complex one.

(2) It is argued that the experimental results reported in this thesis support both the Embodied Appraisal Theory (e.g., Prinz, 2004a, 2004b) and the Social Functionalist Approach to Emotions (e.g., Keltner et al. 2006) that claim human emotions are both biologically based and socially constructed. The results do

not provide a good match to theories such as the evolutionary approach to emotions (e.g., Ekman et al. 1969) that claims that emotions are just biological adaptations, and the social constructionist approach (e.g., Russell, 2003) that claims emotions are just socially constructed. In the experiments reported here, the FACS (Facial Action Coding System) (Ekman and Friesen, 2002) which is a product of Evolutionary Theories, was adapted to create recognizable sequences of facial expressions for the robot head. If emotions were just socially constructed, then a reasonably good recognition accuracy of the sequences of robotic facial expressions could not have been achieved; if emotions were just biological adaptations, then contextual effects of certain surrounding contexts on the robot face would not have been observed. Consequently, the contextual effects found on the robotic facial expressions based on the FACS indirectly support the claims that human emotions are both biologically based and socially constructed.

(3) From a robotic perspective, this thesis should help robotic researchers to build computational models of emotion for socially interactive robots, so that these models can take into account the effects of surrounding emotional contexts in the external environment on people's emotional judgments on the synthetic robot facial expressions. It expands previous investigations by other researchers that have studied the recognition of the facial expressions of the robots in neutral emotional contexts. In general, this research yields significant implications for the designers of social and emotional robots. It seems that the match between a robot's simulated emotions and the surrounding context is important and that the interpretation of a robot's emotional expressions can be affected by what is happening around them. Hong et al. (2002) used neural networks to map real-time surrounding emotional contexts to synthetic avatar emotions – similar approaches in which the valence of surrounding context is detected, and the robot's expressions adapted to match it, may well prove to be a good way of creating robots in the future that are seen as more convincing and believable.

A few papers have also been published during the course of the PhD. The following is the list of the relevant publications:

Conferences:

[1] Zhang, J., and Sharkey, A.J.C.: Contextual Recognition of Robot Emotions. In R. Groß et al. (Eds.) TAROS 2011, LNAI 6856, pp. 78-89. (2011a)

[2] Zhang, J., and Sharkey, A.J.C.: Listening to Sad Music While Seeing a Happy Robot Face. In B. Mutlu et al. (Eds.) ICSR 2011, LNAI 7072, pp. 173-182. (2011b)

Journals:

[1] Zhang, J., and Sharkey, A.J.C.: It's Not All Written on the Robot's Face. Robotics and Autonomous Systems, Volume 60, Issue 11, Pages 1449–1456, DOI: http://dx.doi.org/10.1016/j.robot.2012.05.017. (2012)

Posters:

[1] Zhang, J., and Sharkey, A.J.C.: An Investigation of How Users Perceive the Facial Expressions of an Emotional Robot. Poster presented at the research away day, Sheffield University, Sheffield, UK. (2009)

[2] Zhang, J., and Sharkey, A.J.C.: Contextual Recognition of Robot Emotions. Poster presented at the research away day, Sheffield University, Sheffield, UK. (2011)

1.5 Organization of the Thesis

The thesis has been subdivided into seven chapters: the first one of which is the introduction. Chapter 2 reviews the three classical theories of emotion, and other novel approaches to emotion. Insights from the psychology of emotion have been taken to design the computational models of emotion for some socially interactive robots. In chapter 3 a detailed discussion related to creating believable and convincing synthetic facial expressions of robots has been made.

Examples of socially interactive robots that different computational models of emotion are applied to are listed. Chapter 4 reports a series of 11 experiments that aim to investigate how four different surrounding or preceding contexts (i.e., congruent or incongruent recorded BBC News, selected affective pictures, classical music and film clips) influence users' perception of a robot's simulated emotional expressions. Chapter 4 also discusses the circumstances under which the contextual effects are likely to occur, and analyzes the relative weight of the surrounding context and the robot face on the emotional judgments of the robot face for different contexts. In chapter 5, a further experiment (conducted after the original 11 experiments) that examines how subjects rate the emotional valence of the robot's simulated emotions and each emotional material on its own, is reported. Furthermore, a detailed comparison between the recognition rate and the emotional valence of the robot's simulated emotions and each context is undertaken in Chapter 5 in order to understand the contextual effects found in Chapter 4. Chapter 5 also analyzes the possible factors contributing to the contextual effects found in the original 11 experiments. Based on the analysis, four new experiments including a preliminary experiment designed to identify appropriate contextual materials, and then three further experiments in which factors likely to affect a context effect are controlled one by one, are proposed in Chapter 6. Finally, chapter 7 concludes the thesis with discussions about possible future directions.

Chapter 2: Psychology of Emotion

To date, more and more robots are equipped with synthetic emotions, to communicate with humans. Before wondering if these robots can show us believable and convincing emotions, it is logical to ask, for human beings, what is an emotion? Answering this will not only help people to understand more about themselves as human beings, but also to interact more efficiently with these robots whose emotions are created by human based on what human can understand about emotions. Yet over centuries of debating among philosophers, creative artists, physiologists, animal behaviorists, anthropologists, and sociologists, there is little agreement on what an emotion really is. To lend us support for better understanding of the psychology of emotion with less controversial meaning, Frijda & Scherer (2009) argued that - emotion can be considered as a *bounded episode* in the life of a system that is characterized as an *emergent pattern* of *component synchronization*, preparing *adaptive action tendencies* to *relevant events*, as defined by their *behavioral meaning* and aiming at establishing *control precedence* over behavior. According to them, different theoretical models of emotion differ in their using the highlighted features as italicized in the above sentence. However, in this chapter, the psychology of emotion will not be reviewed based on how different theoretical models of emotion use the highlighted features, but based on the following questions about emotion: (1) what is an emotion? (2) how is an emotion generated? (3) why do humans and other species have emotions?

2.1 Classical Theories

2.1.1 Evolutionary Theories

The evolutionary approach to emotions, assumes that emotions are biologically based and that they are adaptations. The evolutionary approach began with Darwin's (1872/1998) original speculations about emotional expression, defended by James (1884) and Lange (1885) who firstly observed the connection between emotions and the body. It was modernized by Tomkins (1962; 1963) who conceptualized a small, fixed number of discrete ("basic") emotions. Ekman developed it further, and claimed that the Big Six emotions (happiness, sadness, fear, surprise, anger and disgust) are universal and biologically basic (Ekman et al. 1969). He later extended this small set of basic emotions to accommodate many emotions (Ekman, 1999). The evolutionary approach was also updated by Cosmide and Tooby who viewed all of human nature as having evolved through natural selection to successfully solve adaptive problems (Cosmide and Tooby, 1987) and who argued that emotions are genetically coded programs (Cosmide and Tooby, 2000).

In 1872 Charles Darwin published the *The Expression of the Emotions in Man and Animals*, a book that would have a profound impact upon not only emotion research but also the understanding of human origins (Keltner et al. 2006) and that Ekman (2009) has suggested should be considered the book that began the science of psychology. In Darwin's time, facial expressions were thought of as a universal, God-given language created only for humans but not for "lower" species (Bell, 1806; Duchenne, 1862/1990). After observing cross-species and cross-cultural similarities in expressions of emotion, Darwin bolstered his arguments for the substitution of natural selection for God. Ekman (2009) identified Darwin's five main claims about emotional expressions as follows:

(1) Darwin considered emotions as separate discrete entities although dimensions exist to describe intensity or acceptability within each emotion. (2) Darwin focused mainly on facial expression which has been found to be the richest source of information about emotions to date, and spared limited attention to vocalizations, tears and posture. (3) Darwin proposed that facial expressions of emotions are universal but gestures are culture-specific

conventions. The idea that gestures are culture-specific has been shown to be correct by modern psychologists (see Ekman 1976 and ch. 4 in Ekman 1985). (4) Darwin observed that emotions are found not only in humans but also in many other species. (5) Darwin suggested that facial expressions of emotions across various species served functions that he called “serviceable habits”, even if functions may not remain the same as in the past. For example, the raised upper lip that is one of the anger expressions, was explained as exposing canine teeth that can threaten harm to come as well as prepare for the attack. Because in the evolutionary past the facial gesture was serviceable, human beings still make this gesture when they become angry with someone or something.

The influence of the evolutionary approach developed by Darwin on the psychology of emotion has taken three different directions. First of all, it has been argued that emotions are involved for their adaptive value in appraising the fundamental life tasks resulting in problem-solving behaviors that have been adaptive in the past (Ekman, 1999). Secondly, it has been argued that emotions are associated with basic bodily responses and ancient brain structures (Prinz, 2004b), as they involve function through natural selection to increase the survival chance of the gene by recurring reproductive opportunities and suppressing threats to reproduction (see section “Theories of Emotion” in Niedenthal et al. (2006a)’s book “The Psychology of Emotion: Interpersonal Experiential, and Cognitive Approaches”). Thirdly, it has been argued that as a result of adaptive problem-solving behaviors and biological adaptations, certain emotions are universal and biologically basic (Ekman & Friesen, 1971; Ekman et al., 1987; Izard, 1971).

Presumably, emotions have evolved to deal with adaptive problems in the fundamental life tasks (Ekman, 1999). What presumably elicits the emotion (now) and designed the emotion (in the past) consists of the adaptive problems (Niedenthal et al. 2006a). What has been adaptive in the past history of human species partially defines the past, and what has been adaptive in human own individual life history also contributes to the past (Ekman, 1999). Some evolutionary psychologists lend support to the idea of the adaptiveness of emotions for fundamental life tasks. Each emotion is goal-relevant, which prompts us in a direction that can provide better solutions for escaping from universal human predicaments such as losses and frustrations in the course of evolution (Johnson-Laird and Oatley, 1992). For instance, in happiness a goal is attained or maintained, in sadness there is a failure to attain or maintain a goal (Stein and Trabasso, 1992). In other words, human's appraisal of a current event is influenced by their ancestral past (Tooby and Cosmides, 1990; Ekman, 1999).

Human's appraisal of a current event is the product of evolution, resulting in problem-solving behaviors which are also the products of evolution. For example, fear, one of the Big Six emotions, is said to have evolved to cope with dangers (Plutchik, 1980). Fear can result in behaviors of fleeing or fighting that allow us to cope with potential threats which pose major survival challenges effectively. Such psychological mechanisms are the kind of thing that evolution would have selected for. Thus, fear is the product of evolution. Moreover, anger is associated with enhanced distribution of blood to the hands to fight an enemy, whereas fear involves less blood flow to the periphery to escape an attack with minimal loss of blood, if necessary in both situations (Levenson, 1992). Such problem-solving behaviors associated with a small set of emotions can be listed. There are eight adaptive behaviors, namely, withdrawing, attacking, mating, crying for help, pair bonding, vomiting, examining, and stopping/freezing, and they are associated with fear, anger, joy, sadness, acceptance, disgust, expectancy, and surprise (Plutchik, 1984). In another version proposed by Maclean (1993), six behaviors - searching, aggressive, protective, dejected, gratulant (triumphant), and caressive, have been linked to six corresponding basic emotional expressions - desire, anger, fear, dejection (sadness), joy, and affection, respectively. All these behaviors are related to opportunities and threats to reproduction (Niedenthal et al. 2006a). Thus, they are also the products of evolution.

As can be seen, natural selection has a heavy impact on appraisals in the fundamental life tasks and problem-solving behaviors in the distant past, but what lies at the heart of the evolutionary approach is that emotions are biological adaptations from the distant past. First of all, human emotions are biologically based. James (1884) and Lange (1885) firstly observed the connection between emotions and the body. According to them, an emotion is a perception of a patterned change in the body. For instance, people feel fearful as sometimes they perceive their hearts racing, their lungs inhaling, and their muscles tensing. Secondly, emotions are associated with ancient brain structures. More recently, two imaging studies of the emotions have shown that guilt and love are associated with bodily perturbations. When subjects recalled episodes of guilt, activation in anterior cingualte cortex and the insula was found (Shin et al. 2000), and when subjects viewed photographs of their lovers similar activations were found (Bartels and Zeki, 2000). All these studies show that the brain structures underlying emotion are associated with the perception and production of bodily response (Damasio, 1999). Cosmides and Tooby favored Darwin's assumption of emotions evolved function through natural selection, to update the link between emotions and the body. According to their evolutionary thinking, emotions are genetically coded programs as human brain structure and computational architecture of mind are the products of evolution to successfully solve adaptive problems through natural selection (Cosmides and Tooby, 1987). And such genetically coded programs are firstly triggered by evolutionarily recognizable biologically-relevant objects or adaptive problems, then a number of the body's functions (perception, and physiological reactions etc.) are coordinated to solve adaptive problems (Cosmides and Tooby, 2000). To sum up, evidence that emotions are biological adaptations in the distant past strongly supports the evolutionary approach as it shows that emotions are genetically-orientated and thus are hard-wired.

Since emotions have evolved for adaptive appraisals in fundamental life tasks resulting in adaptive problem-solving behaviors, and they are biological adaptations in the distant past, it is logically consistent to expect that certain emotions are basic and universal, as they have evolved for certain functional, adaptive reasons. For instance, they serve to convey messages about the expresser (e.g., communicative functions) as well as to experience the emotions in the receiver's body (perceptive functions). Darwin considered emotions as separate discrete entities, or modules, such as anger, fear, etc (Ekman, 2009). Tomkins treated each basic emotion as a single brain process (an "affect program") which can vary in intensity, which when it is triggered, produces all the various components of the emotion (i.e., facial and vocal expression, changes in peripheral physiology, varying from subjective experience, instrumental action) (Russell et al. 2003). However, Tomkins argued that the perception of changes in the face provides more specific feelings of emotion than that of other part of the body. Moreover, the categories of emotion were thought to correspond to specific categories of facial expressions, through the actions organized around their facial expressions. For example, emotions related to disgust derive from the prototype of rejecting food that is noxious or dangerous to eat, with a core expression of opening the mouth and lips, and pushing out with the tongue. This prototype disgust reaction has generalized to other rejection scenarios, such as the emotion of contempt, where the object is another person, and the emotion of shame, where the object is the self. Another distinctive aspect of his theory is accounting for the way in which the same message is encoded and decoded: If (except in cases of deliberate, socially induced deception) expressions of emotion broadcast veridical information that the receiver recognizes, then researchers can focus on either the encoding or decoding side (Russell et al. 2003).

Consistent with Darwin's original speculations about emotional expression, Tomkins's findings that there are discrete and basic emotions which are mainly perceived through face, and that the same message signaled by the facial expression is encoded and decoded, was largely responsible for encouraging the work of Ekman in the late 20th century. Initially, Ekman wanted to examine whether expressions of Tomkins's basic emotions are universal and relied on decoding (judgments by observers) to establish just what emotion a specific expression of emotion represents.

In 1972, Ekman used four experiments co-conducted by his colleague Friesen, to demonstrate that there are universal facial expressions of emotion (Ekman, 1972). Data on five literate cultures, namely four Western and one Eastern, and on two preliterate cultures from New Guinea, were reported. The samples were drawn from six different language groups: Dani, English, Japanese, Fore, Portuguese and Spanish. In the first experiment, Japanese and American observers were asked to judge whether particular spontaneous facial expressions of Japanese and American subjects, who were videotaped while watching a stressful or non stressful film, have occurred. It was shown that, the facial expressions of the Americans were interpreted in the same way by the Japanese and American observers, as were the facial expressions of Japanese. In the second experiment, when exposed to a stressful film, the same specific facial behaviors but not necessarily the same specific emotional expressions, were shown in both Japanese and American culture. Furthermore, in the third experiment, over 30 still photographs of humans' facial expressions selected from 3000 samples were presented to five literature cultures (Chile, Argentina, Brazil, Japan and the United States). In the experiment, the observers were asked to choose from among six emotion categories (happiness, fear, surprise, anger, disgust/contempt, and sadness) the one which best described each photograph, and it was shown that six distinct facial expressions are universal (e.g., the Japanese recognized happiness perfectly by 100%, followed by the Argentinean who scored 98%, the Brazilian and Chilean achieved the same rate as 95%, while the American gained 97%, and cross-group recognition of other facial expressions achieved similar but lower rate.). Further evidence of the universality of facial expressions of emotion was obtained in the fourth experiment, in which the same facial expressions are associated with the same emotions in preliterate (i.e., New Guinea) as in literate cultures.

Ekman thus confirmed that both Darwin and Tomkins were right and that the Big Six emotions (happiness, sadness, fear, surprise, anger and disgust) are universal and biologically basic. Based on this conclusion, Ekman and Friesen (1978) developed a Facial Action Coding System (FACS), revised in 2002, to create a reliable means for skilled human scorers to determine the category or categories (include the six distinct facial expressions) in which to fit each facial behavior of humans. It is a comprehensive and anatomically based tool for describing/measuring any of the 44 human facial movements, resulting in its being used by professionals such as psychologists, psychiatrists, and neurologists, worldwide.

In summary, evolutionary theories of human emotions mainly focus on defining what an emotion is.

They emphasize the impact of natural selection on appraisals of fundamental life tasks, problem-solving behaviors in the distant past and biological adaptations, resulting in the conclusion that certain emotions are universal and biologically basic. Evidence that emotions are biological adaptations is central to the evolutionary approach to emotion and it is widely accepted that a small set of emotions are biologically based. Additionally, evolutionary theories of human emotions also provide answers to the question like why do humans and other species have emotions? For instance, Ekman (1999) argued that emotions have evolved to deal with adaptive problems in the fundamental life tasks. However, Prinz (2004b) criticized Ekman's theory (Ekman, 1999) as immodest, as Ekman's basic emotion list was expanded to include: amusement, contempt, contentment, embarrassment, excitement, guilt, pride in achievement, relief, satisfaction, sensory pleasure, and shame. This criticism was made based on the argument that not every emotion is part of human's bioprogram (Prinz, 2004b).

2.1.2 Cognitive-appraisal Theories

A. Appraisal Theories of Emotion

Appraisal theories of emotion, according to Scherer (2009), have roots in Aristotle, Descartes, Spinoza and Hume. They were firstly explicitly formulated by Arnold (1960) and Lazarus (1966, 1991). And in the early 1980s (see the historical reviews by Scherer 1999, 2001), they were actively developed by Ellsworth and Scherer and their students (Scherer 1984, 2001; Smith &

Ellsworth 1985; Roseman & Smith 2001; Sander et al. 2005). Where they differ from the evolutionary theories of emotion that link emotions to biological adaption in the distant past, is that they link them to more immediate cognitive processes of evaluation of meaning, causal attribution, and assessment of coping capabilities (Niedenthal et al. 2006a). Appraisals, according to a definition by Clore and Ortony (2000), are psychological representations of emotional significance for the person experiencing the emotion. And a central tenet of appraisal theory, reviewed by Scherer (1999), is the claim that emotions are elicited and differentiated on the basis of a person's subjective evaluation or appraisal of the personal significance of a situation, object, or event on a number of dimensions or criteria.

Modern cognitive appraisal theories began with Arnold (1960) who first used the term "appraisal" to explain the differentiated emotions. She proposed that the cognitive and evaluative preconditions of emotions are appraised according to three dimensions: evaluation of the object as beneficial or harmful for oneself; presence or absence of the object, and the ease or difficulty to attain or avoid the object. Reisenzein (2006) gave a detailed review of how Arnold's theory of emotion can be viewed as in a structural format:

"For example, according to Arnold, joy is experienced if one believes that an object (a state of affairs) is present, is positive, and "rests in possession" (i.e., can be easily maintained). Sorrow or sadness occurs when a negative state is present but "conditions are favorable," that is, one believes one can cope with the negative state. Fear occurs if one believes that a negative event is absent (not yet present, but a future possibility) and is "too difficult to cope with" (Arnold, 1960a, p. 194). Hope occurs if one believes that a positive future state can be attained. Several further emotions, including pity, guilt, remorse, shame, embarrassment and admiration, are analyzed from an appraisal perspective in Part III of Arnold (1960b). For example, according to Arnold (1960b, p. 318) pity is felt "whenever another's suffering is realized and appraised as bad." Guilt is experienced when one believes one has culpably broken a moral rule (Arnold, 1960b, pp. 291-292). Shame and embarrassment are caused by the appraisal that one is not conforming to internalized ideal or social norms of appropriateness of conduct or appearance (see Arnold, 1960b, p. 299f).” (Reisenzein, 2006, p.930-931)

Arnold (1960) emphasized that the evaluation process is typically “direct, immediate, intuitive” and concluded that:

“We can now define emotion as the felt tendency toward anything intuitively appraised as good (beneficial), or away from anything intuitively appraised as bad (harmful). This attraction or aversion is accompanied by a pattern of physiological changes organized toward approach or withdrawal. The patterns differ for different emotions.” (Arnold, 1960a, p.182)

Richard S. Lazarus (1922-2002), the theorist who was mainly responsible for popularizing the appraisal concept, was significantly influenced by most assumptions of Arnold's theory, such as emotions are object-directed, emotions presuppose cognitions and emotions also presuppose desires (Reisenzein, 2006). Lazarus (1966) identified two different appraisal processes, namely primary appraisal and secondary appraisal, to develop a theory about emotions and stress. Primary appraisal differs from secondary appraisal in terms of sources of information they rely on. In details reviewed by Krohne (2002), primary appraisal concerns whether something of relevance to the individual's well being occurs (three components are distinguished: goal relevant, goal congruence, and type of ego-involvement), whereas secondary appraisal concerns coping options (likewise three components are distinguished: blame or credit, coping potential, and future expectations). Moreover, Lazarus (1966) elaborated the concept of appraisal introduced by Arnold (1960) as processes that allow for re-appraisals of objects or events based on new information or evaluations. Later, Lazarus (1991) distinguished 15 basic emotions. Nine of these are negative (anger, fright, anxiety, guilt, shame, sadness, envy, jealousy, and disgust), whereas four are positive (happiness, pride, relief, and love) (Two more emotions, hope and compassion, have a mixed valence). Consistent with Arnold's theory, Lazarus’s theory shared that at least a set of “basic” emotions have an evolutionary basis (Lazarus, 1991).

Pioneering research done by Arnold (1960) and Lazarus (1966, 1991) led modern appraisal theories to evolve, first independently of each other, but rapidly showing a high degree of convergence (Sander et al. 2005). What lies in the heart of these appraisal theories, is that the nature of an emotional reaction can be best predicted on the basis of individual's subjective appraisal or evaluation of an antecedent situation or event (Scherer, 1999). In spite of the fact that modern appraisal theories vary on the number of dimensions needed to describe basic emotional experience, a limited set of predictor dimensions can generally correctly classify 40% to 50% of 15-30 emotional states in empirical studies (Scherer, 1999). Most of the empirical studies require subjects to infer or recall the nature of their event or situation appraisal, mostly by rating scales or questionnaire items constructed on the basis of theoretically assumed appraisal dimensions. However, the use of self-report of emotion-antecedent appraisal and the process of imaging an emotional event in these studies, have given rise to repeated criticism (Scherer, 1999).

Scherer (1997) chose a set of criteria, which he called stimulus evaluation checks (SECs), to represent the minimal set of dimension or criteria that are considered necessary and sufficient to differentiate the major families of emotional states. The SECs were defined in his 2001's version of component process model of emotion (CPM; see Scherer, 1984a, 1984b, 1986, 1993, 2001). In the framework of the CPM (see Fig. 2.1), emotion is defined as an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism (Scherer, 2001). In light of CPM, Emotion is considered as a theoretical construct that consists of five components corresponding to five distinctive functions (see Table 1).

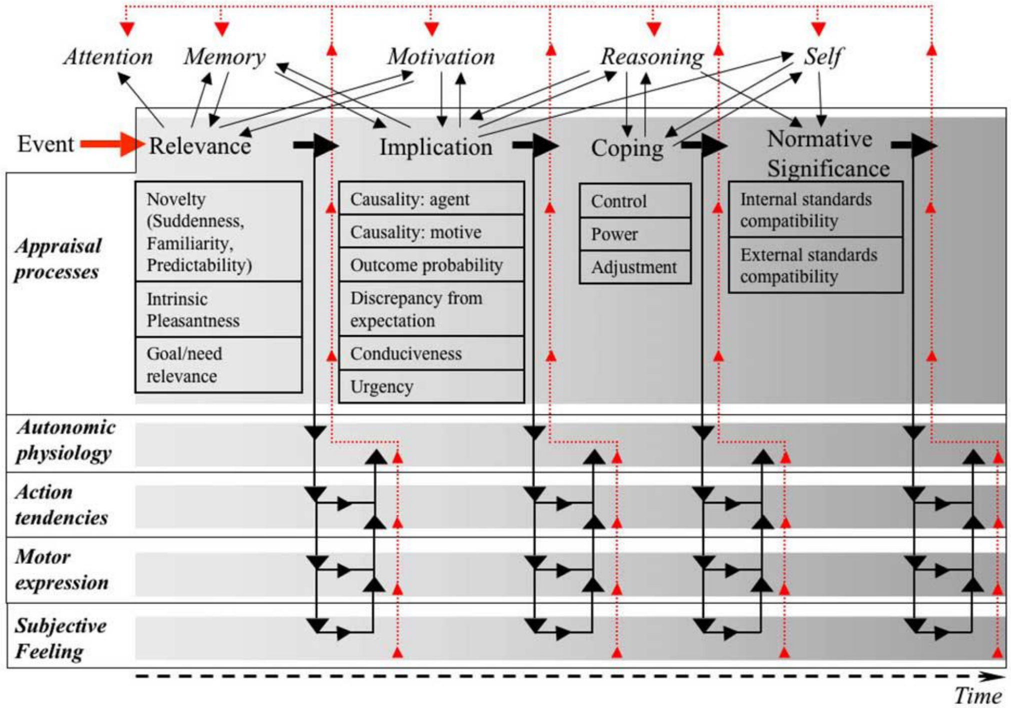


Fig. 2.1: Comprehensive illustration of the CPM of emotion (Sander et al. 2005), permission to reprint granted.

Table 2.1 Relationships between the functions and components of emotion, and the organismic subsystems that subserve them (Sander et al. 2005), permission to reprint granted.

Emotion function	Emotion component	Organismic subsystem (and major substrata)
Evaluation of objects and events	Cognitive component	Information processing (CNS)
System regulation	Peripheral efference component	Support (CNS, NES, ANS)
Preparation and direction of action	Motivational component	Executive (CNS)
Communication of reaction and behavioral intention	Motor expression component	Action (SNS)
Monitoring of internal state and organism-environment interaction	Subjective feeling component	Monitor (CNS)

CNS, central nervous system; NES, neuro-endocrine system; ANS, autonomic nervous system; SNS, somatic nervous system. The organismic subsystems are theoretically postulated functional units or networks.

As can be seen in the CPM, four major appraisal objectives that an organism needs to reach to adaptively react to a salient event are suggested. These objectives are obtained by organisms evaluating the event and its consequences on a number of criteria or stimulus evaluation checks(SECs), with the results reflecting the organism’s subjective assessment of consequences and implications on a background of personal needs, goals and values(which may well be unrealistic or biased) (Scherer, 2001; Sander et al. 2005; Scherer, 2009). According to CPM, SECs are processed in sequence, following a fixed order, consisting of four stages in the appraisal process that corresponds to the appraisal objectives described: (1) relevance (a stimulus event is considered as requiring attention deployment, further information processing and potential action); (2) implications (following attention deployment, the pertinent characteristics of the stimulus event and its implications or consequences for the organism are determined); (3) coping potential (once the nature of event and consequences are known sufficiently well, the organism checks its ability to cope with the consequences to be expected); (4) normative significance (overall assessment of the event with respect to compatibility with self-concept, values, social norms and moral rules) (Scherer, 2009). Moreover, it is postulated by the CPM that external or internal event changes maintain a recursive appraisal process until the monitoring subsystem signals termination of or adjustment to the stimulation that originally elicited the appraisal episode (Scherer, 2001; Sander et al. 2005).

B. Comparing Evolutionary and Appraisal Accounts

Inspired by the observation that different individuals can experience different emotions in response to the same event or stimulus, the CPM was initially designed to explain what caused an emotion to occur in the first place which the evolutionary approach failed to. It is then logical to ask, can the model predict different emotions? The CPM makes specific predictions about the effects of the results of certain appraisal checks on the autonomic and somatic nervous systems, indicating which somatovisceral changes and which motor expression features are expected (Scherer, 2009). In CPM, emotion differentiation is the result of the net effect of all subsystem changes brought about by the outcome profile of the SEC sequence (Scherer, 2001; Sander et al. 2005; Scherer, 2009). These subsystem changes are theoretically predicted on the basis of a componential patterning model, the central assumption of which is that the different organismic subsystems are highly interdependent and that changes in one subsystem will tend to elicit related changes in other subsystems.

As illustrated in Fig. 1, the CPM process is recursive and each SEC is consecutive, resulting in specific motivational and behavioral tendencies which are expected to be activated in the executive subsystem in order to serve the specific requirements for the adaptive response demanded by a particular SEC result (Scherer, 2001; Sander et al. 2005; Scherer, 2009). This is consistent with the claim in evolutionary theories that there are some major patterns of adaptation in the life of animate organisms that reflect frequently recurring patterns of environmental evaluation and adaptation results (Scherer, 2001; Sander et al. 2005). However, the assumption of a limited number of innate, hard-wired affect programs for basic emotions such as anger, fear, joy, sadness, and disgust, in evolutionary theories, is not endorsed in the CPM. In contrast, the CPM bolsters the argument that an extraordinarily large number of different emotions can be yielded, virtually as many as there are different integrations of appraisal results and consequent response patterns, as the emotion process is a continuously fluctuating pattern of change in several organismic subsystems (Scherer, 2001; Sander et al. 2005). Based on this argument, the Geneva Emotion Wheel was designed (see the first version in (Baenziger et al., 2005)) and developed (see the second version in (Scherer, 2005)). The wheel is a circular arrangement of 20 emotion families arranged in two-dimensional space. And on the basis of the distance from the hub of the wheel and the size of the circles, the intensity of an experienced or imagined emotion can be rated. Basically, the wheel is an experimental tool to measure subjectively experienced feeling.

It seems that appraisal theories of emotion can account for the emotions of people experiencing feelings, as their feelings can be measured by the Geneva Emotion Wheel. Additionally, appraisal theories of emotion can also predict the emotions of the person experiencing the feelings. For instance, GENESE—Geneva expert system on emotion, can formalize the appraisal predictions in the form of a simple expert system, and subject the system to empirical test. Scherer (1993) reported that the GENESE obtained an accuracy percentage of 78 percent for the expert system's diagnoses, with over 200 emotion situations reported by different subjects. In the empirical test, subjects were asked to think of a situation in which they experienced a strong emotion, and their emotions were diagnosed by the GENESE on the basis of the responses to questions that represent simplified versions of the SECs. An improved version of GENESE is available for testing on the web (www.affective-sciences.org/genese).

In summary, appraisal theories of emotion do not totally reject the claim in evolutionary theories that emotions are adaptations, but they are better than evolutionary theories in providing explanations about what caused an emotion to occur in the first place. Therefore, appraisal theories of emotion mainly investigate how an emotion is generated. At the same time, they also contribute to the definition of emotion, as their products such as Geneva Emotion Wheel and GENESE can be used to measure and diagnose emotions of the person experiencing the feelings. In addition, according to Niedenthal et al. (2006a)' review of appraisal theories of emotion, during the evaluation or appraisal of the personal significance of a situation, or an object, different physiological process and facial expressions are produced, but the antecedent of the emotion – the specific profile of appraisal – determines which discrete emotion is experienced. Consequently, appraisal theories predict that different individuals will experience different emotions in response to the same event or stimulus.

However, cross-cultural differences were found in the use of appraisal dimensions (Scherer, 1999). And all current appraisal theories have been criticized for their predominantly intrapsychic orientation, tending largely to neglect the social context in which the emotion-antecedent appraisal and the ensuring response are often embedded (Kappas, 1996; Parkinson, 1997).

2.1.3 Social Constructionist Theories

As a general rule, the social constructionist approach firstly rejects the idea that emotions are biological adaptations, secondly it embraces the evidence that emotions vary across cultural boundaries, and finally it arrives at the conclusion that emotions are socially constructed.

According to this approach, emotions can be learned and thus are not innate. This perspective has its root in the work of Averill (1980) which defined emotion is as “a transitory social role (a socially constituted syndrome) that includes an individual’s appraisal of the situation and that is interpreted as a passion rather an action.” His work highlighted the important role of learning in human emotions. This perspective is also supported by Armon-Jones (1986) who suggests that an individual cannot experience any given emotion until he/she has learned to interpret situations in terms of the standards and moral imperatives endorsed by the culture that are relevant to the emotion.

Pollak and his colleagues continued to bolster the claim that perceptual learning plays an important role in emotional development. They demonstrated this perspective in three experiments. In the first experiment, it was found that although young children perceive emotional expressions based upon similar categories, abused children–who have experienced frequent displays of extreme hostility–display broader perceptual categories of anger (but not other emotions) relative to nonmaltreated children (Pollak & Kistler, 2002). It was observed in the second experiment that abused children could accurately identify angry faces based upon less perceptual information than typically developing children (Pollak & Sinha, 2002). Their third experiment pushed the examination of the role of perceptual learning in emotional development a bit further (Pollak et al. 2009). In the experiment, a group of abused children who had been exposed to extremely high levels of parental anger expression and physical threat were tested. Children were presented with arrays of stimuli that depicted the unfolding of facial expressions, from neutrality to peak emotions. It was found that the abused children accurately recognized anger early in the formation of the facial expression, when few physiological cues were available. Moreover, the speed of children’s recognition was associated with the degree of anger/hostility reported by the child’s parent. Their findings indicated that processing and recognition of emotional expressions can be affected by the social relevance of facial movements.

The second strategy for social constructionist approach is to emphasize that emotions are cultural adaptations rather than biological adaptations. In the evolutionary approach, classical studies used photographs depicting nothing other than people’s posed facial expressions which had the effect of boosting the capabilities of observers across cultures to infer the emotional state of those depicted. However, in real-life situations, facial expressions occur in contexts that have meaning beyond the expressions themselves. In such situations, Russell et al. (2003) argued that attributing a specific emotion to the sender becomes more complex, as even when the stimuli are facial expressions of basic emotions, the context which they occur (Carroll and Russell, 1996), the receiver’s current affective state (Niedenthal et al. 2000), and the gender of the sender (Widen and Russell, 2002), can color the attribution. In fact, there was a growing body of cross-culture studies of facial expressions (Russell, 1991; Russell, 1994; Haidt and Keltner, 1999; Elfenbein and Ambady, 2002) and vocal expressions (Scherer et al. 2001) challenging the universality of basic emotions. As an illustration, in a study by Russell (1994), recognition scores for six facial expressions of emotion were compared between three different culture groups, namely Western Literate, Non-Western Literate and Isolated Illiterate (uneducated persons). It was indicated that although receivers from different cultures agree better than chance on the best label to assign to posed, static facial photographs, the recognition scores are proportional to the amount of Western influence which may result from a force-choice response format. More recently, Naab and Russell (2007) used spontaneous facial expressions of New Guineans to indicate that recognition was typically lower than that of posed photographs. These studies showed that the similarity between the sender’s and the receiver’s language and culture could color the attribution of specific emotions (Russell et al. 2003). Even Ekman himself (1999) had to admit that the relationship between a facial configuration and what it signified is socially learned and culturally variable, and he talked about universal emotion families, rather than universal emotions, indicating that cultures may customize human’s innate affective stock in different ways. Therefore, Russell (2009) suggested the replacement of Ekman’s universality of basic emotions by Minimal Universality.

As firstly defined in (Russell and Fern ́andez-Dols, 1997b), Minimal Universality predicts something of cross-cultural similarity in interpreting facial expressions without postulating an innate emotion signaling system.

The third strategy mainly concerns methodologies such as categorizing and labeling emotions in a social construction way. Developments in this approach, as reviewed by Scherer (2009a), were based on James (1890; 'perception of bodily changes is the emotion'), and modified by Schachter & Singer (1962; 'perceived arousal leads to labeling feelings as an emotion based on situational cues'). They were revived by Russell (2003; 'continuous core affect—constituted by valence and arousal is interpreted and categorized in the light of situational cues') and Barrett (2006; 'core affect is differentiated by a conceptual act that is driven by embodied representations and available concepts'). Since there was cross-cultural difference in attributing specific emotions to facial expressions, Russell (2009) also suggested the replacement of discrete emotions by Core Affect. Scarantino (2009) described "core affect" as follows:

"Core affect, understood as the category comprising the set of all possible valence and arousal combinations on the circumplex, differs from discrete emotions in three crucial ways: it is ubiquitous, it is objectless, and it is primitive." (Scarantino, 2009, p. 948)

As can be seen, on the one hand the social constructionist view shares two common features with the cognitive appraisal approach. One is the claim that emotions are a result of an assessment of the situation and one's capacity to deal with it, the other is to deny that biologically relevant antecedents completely determine the emotion. On the other hand, the social constructionist approach differs from the cognitive appraisal approach in that the moral values of the culture provide the specific content to the appraisal and meaning to the situation. Consequently, similar to the cognitive appraisal approach, the social constructionist approach also considers how an emotion is generated, but it mainly focuses on defining what an emotion is, e.g., Russell (2009) suggested that the discrete emotions should be replaced by Core Affect.

However, critics such as Prinz (2004b) argued that firstly constructionists may be mistaken to conjecture that emotions can be disembodied, and secondly, that constructionists over emphasized the role of cognitions in emotion. Therefore, constructionists such as Barrett (2006) made a concession that the importance of evolutionarily preserved responses should not be denied, but any privileged status of emotions treated as innate neural circuits or modules should be denied.

2.2 Novel Approaches

The three traditional theories of emotion have been reviewed above. Generally, Social Constructionist Theories are not as popular as Evolutionary Theories or Cognitive-appraisal Theories. However, Moors (2009) pointed out that there are other approaches to emotion that differ from the three classical theories because they focus on a specific aspect or component of emotion, such as motivation or action, preparation, or combining features from the three major orientations. In other words, while the three classical theories lend us support in a basic understanding of the definition of emotions, there are approaches that combine insights from these three classical theories, to provide us with more explanation of the generation and the purpose of emotions. As a result, other approaches, such as those described below, seem to be more novel, compared with the three traditional theories.

2.2.1 Theory of Embodying Emotion

Emotional cues can be facial expression features, words or phrases that contain certain kinds of emotional information, musical rhythm or melody that can elicit emotions, odors that can stimulate olfactory system, and so on. People experience emotional feelings themselves and perceive emotions of others via single or multiple emotional cues. But how is the emotional information of these emotional cues processed?

When human brains process emotional information from different modalities, it involves highly overlapping mental processes (Niedenthal et al. 2005; Damasio, 1994), which results in modalities fusion. In other words, populations of neurons in the modality-specific sensory, motor, and affective systems are highly interconnected, and their activation supports the integrated, multimodal experience of different emotional states (Niedenthal et al. 2007). However, only a partial multimodal reenactment of the experience is produced (Damasio, 1989; Barsalou et al. 2003), as attention is selectively focused on the aspects of the experience that are most salient and important for the individual, and are most likely to be stored for later reactivation (Barsalou, 2003).

According to Niedenthal's theory of embodying emotion (Niedenthal et al. 2007), circuits exist which can be fast, refined, and able to flexibly process many emotional states that can be reactivated without their output being observable in overt behaviour, in human brains' modality-specific systems. The reenactment of the other's emotional states or emotional states in some specified situations can be sometimes accompanied by self-reports of the associated emotional state (Damasio, 1994) as this is part of the bodily re-enactment. It can be measured by a technique called electronmyographic recording.

The reenactment of the emotional states after receiving the emotional cues via a certain modality may, but not definitely will, affect human personal emotional states before the reenactment to a certain degree. It can be measured by Affect Grid (Russell et al. 1989), Brief Mood Introspection Scale (Mayer and Gaschke, 1988), PANAS (Watson et al. 1989), and so on.

Emotion congruence in perception strongly affects later reactivation and emotion recognition. First of all, shifting from processing in one modality to another involves temporal processing costs (Spence et al. 2001). Secondly, individuals simulate objects in the relevant modalities when they use them in thought and language because the more relevant the modalities are, the less processing costs will be produced (Pecher et al. 2003). Moreover, for positive and negative concepts, verifying properties from different modalities produced costs such that reaction times were longer and error rates were higher than if no modality switching was required (Vermeulen et al 2007). Consequently, on one hand, emotion congruence in perception facilitates later reactivation, whilst emotion incongruence in perception impairs later reactivation; on the other hand, emotion congruence in perception shortens the reaction times and minimizes the recognition errors, whilst emotion incongruence in perception increases the reaction times and maximizes the recognition errors.

It was reported in the neuroimaging studies that recognizing someone else's facial expression of an emotion or experiencing that emotion oneself involves the reactivation of these neural circuits (Niedenthal et al. 2007). Furthermore, it was also demonstrated that using emotional information stored in memory involves embodiment and embodiment does not occur when the information can be processed on the basis of association or perceptual features (Niedenthal, 2009).

For instance, when someone observes a facial expression of an emotion in another person, he/she firstly simulates the emotion in his/her brain supported by the promising theory of "mirror neuron system". Then the other person's facial expression is imitated by the observer accompanied by self-reports of the associated emotional state (Chartrand and Bargh, 1999) as this is part of the bodily re-enactment of the experience of the other's state (Niedenthal et al. 2007). Finally, the emotion can be recognised. However, if someone's motor movements are inhibited, then there will be interference in the experience of emotion and processing of emotion information (Niedenthal et al. 2005). By contrast, if the imitation goes smoothly, a strong foundation for empathy or even marriage can be laid (Decety and Jackson, 2004).

To sum up, the theory of embodying emotion shares one perspective with evolutionary approach that human emotions are biologically based, as has been defended by James (1884) and Lange (1885)'s theory for over more than one century. James and Lange associated emotions with perceptions of bodily changes. Antonio Damasio (1994) enriched James and Lange's legacy by the claim that the brain centers associated with emotion are also associated with perception and regulation of the body. The theory of embodying emotion has continued to gather evidence from neurobiological and imaging studies for decades. It may provide an explanation of emotional congruence in perception of human emotions and of why if body perception is impaired emotions wane. Nevertheless, it has provided us a better understanding of how an emotion is generated, and of how an observer recognizes a facial expression of an emotion in another person.

2.2.2 Embodied Appraisal Theory

The embodied appraisal theory was firstly presented in Prinz's book (2004a) called *Gut Reactions: A Perceptual Theory of Emotion*, where it is argued that emotions are perceptions of bodily changes that represent such things as dangers, losses, and offenses, because they are set up to be set off by such things. He explained the reasons for the theory's name: firstly, emotions are embodied because they are perceptions of bodily changes; secondly, they are appraisals because they represent matters of concern. James (1884) and Lange (1885)'s theory that associated emotions with perceptions of bodily changes, Lazarus (1991)'s taxonomy of core relational themes —e.g., dangers for fear, losses for sadness, insults for anger, and so on, and Damasio's (1999) theory that the brain structures underlying emotion are associated with the perception and production of bodily response, were mainly responsible for the creating of the embodied appraisal theory.

The following content from his paper (Prinz, 2004b) explained how his theory was inspired by the above three theories:

> "Consider how this works in a typical case. You hear a loud sudden noise. That auditory state sets your body into a patterned response. The response is perceived. The perception of that [sic] your response is your fear. Loud noises are not the only fear trigger, however. A sudden loss of support, a snake, or an infelicitous election return can all have the same impact. Our mental representations of all these fear elicitors group together into a mental file. When any item in the elicitation file is activated, fear results. Fear represents danger in virtue of the fact that, collectively, the items in the elicitation file calibrate fear to danger, and they have the function of doing so. Danger is what unites all the disparate contents of the elicitation file. Fear represents danger because it has the function of occurring when danger occurs, and it obtains that function via an elicitation file filled with a wide range of perceptions and judgments." (Prinz, 2004b, p. 14)

As can be seen, James and Lange's theory was only adapted in the embodied appraisal theory under the condition that perceptions of patterned changes in human body are set up to be set off by an elicitation file. The elicitation file stores a highly plausible list of formal objects that presented by the core relational themes of emotions outlined by Lazarus. For example, Lazarus identifies "a demeaning offense against me and mine" for anger and "facing an immediate, concrete, and overwhelming physical danger" for fright.

In the embodied appraisal theory, according to Prinz (2004b), emotions can be both embodied and socially constructed. Before compromising between evolutionary approach and social constructionist approach, he criticized both approaches (Prinz, 2004b). For the evolutionary approach, firstly cross-cultural research on facial expressions did not show that certain emotions are universal, secondly, to show that an emotion is adaptive does not entail that it is a biological adaptation. For social constructionist approach, firstly assuming that emotions can be disembodied may be mistaken, secondly, constructionists over emphasized the role of cognitions in emotion. The embodied appraisal theory, according to Prinz (2004b), can reconcile these problems. It is because emotions are simple perceptions of bodily changes, and they carry information by being calibrated through elicitation files to matters of concern. Culture can not only inform elicitation files, but also alter human patterns of bodily response. Therefore, his theory can allow for a range of different kinds of perceived body states, from universal "innate" ones to culturally specific learned ones, to count as emotions.

To sum up, similar to the theory of embodying emotion, the embodied appraisal theory mainly focuses on the generation of an emotion of the person who is experiencing it. However, according to book reviewer Craig DeLancey on *Gut Reactions: A Perceptual Theory of Emotion*, Prinz has no compelling evidence that emotions are embodied appraisals, which is probably the biggest weakness of the embodied appraisal theory. In spite of this, the theory itself provides inspiring insight that every emotion that is given a name is the product of both nature and nurture.

2.2.3 Social Functionalist Approach

So far, as reviewed above, different approaches to emotion mainly focused on two aspects of emotions, one is what an emotion is, the other is how emotions are generated, meaning that little was known about why do humans or other species have emotions. This question of “why” has driven many scientists to investigate emotions through a social functionalist approach. The social functionalist approach to emotion has roots in Charles Darwin’s book (1872/1998), *The Expression of the Emotions in Man and Animals,* which suggested that facial expressions of emotions across various species served functions that he called “serviceable habits”, even if those functions may not remain the same as in the past (Ekman, 2009). Modern social functionalists, such as Dacher Keltner, followed Darwin’s original speculations about emotional expression, to investigate emotions in a more systematic way. More recently, Keltner et al. (2006) concluded that emotions serve survival and reproductive functions that are best understood at four levels simultaneously – intra-individual (i.e., inform individual of problems/opportunities, and prepare the individual for action), dyadic(i.e., knowledge of others' mental state, reward or punish prior action, and evoke complementary or reciprocal behavior), group (i.e., define group boundaries and members, define group roles and identities, and motivate collective action), and cultural (i.e., define cultural identity, identify norms and values, and reify cultural ideologies and power structures). Whereas functions at the individual level typically involve basic survival, functions at the other levels involve the facilitation of social bonding and collaboration.

The social functionalist approach to emotions has been developed for more than one century. It combined insights generated by the evolutionary approach to emotions and the social constructionist approach. There are three insights generated by evolutionary approach to emotions, as reviewed by Keltner et al. (2006): (1) emotions enable the individual to respond effectively to environmental challenges and opportunities; (2) emotions enhance survival and reproduction rates of the individual, offspring, and related kin, given the physical and social conditions of the Environment of Evolutionary Adaptedness (EEA; Tooby & Cosmides, 1990); (3) emotions enable social commitments (e.g., to pair bonding, to parent-child bonds, to cooperative alliances, or to group memberships) which help to solve problems of reproduction and the problem of cooperation. The social functionalist approach also generated three insights (Keltner et al. 2006): (1) many emotions act as social-moral intuitions, for examples, the experience of anger correlates with judgments of violated rights, disgust with violations of purity, contempt with violations of duties and obligations, and sympathy with perception of harm to others (Eisenberg, et al., 1989; Haidt, 2003; Rozin et al., 1999; Vasquez, et al., 2001); (2) emotional communication coordinates social interactions; (3) many individual differences in personality reflect differences in emotionality.

The combination of insights generated by evolutionary approach to emotions and social functionalist approach gave birth to recent theory of emotions presented in (Keltner et al. 2006). Keltner et al. arrived at the conclusions that, firstly, the functions of emotion depend on the four levels of analysis, as mentioned above; secondly, the four levels of analysis reconciles the conflicting conclusions of evolutionary theorists and social constructionists; and finally, the functions of emotions make human different from other primates. They argued that at the individual and dyadic levels of analysis, cross-cultural research on emotion supports a universalist position whereas cultural differences in the practice of emotion move to the group and cultural levels of analysis. Moreover, because human emotions are shaped by language, humans have much larger frontal cortexes than that of other primates, and human have capacity for culture, humans are different from other primates.

To summarize, the social functionalist approach to emotion, especially Keltner et al. (2006)'s four levels of analysis, provides answers as to why human emotion is unique, and why both evolutionary theorists and social constructionists are right about their claims in cross-cultural studies.

2.2.4 Perceptual Control Theory on Emotions

The historical development of ideas pertaining to Perceptual Control Theory (PCT), can be traced back to (Dewey, 1896) in which the idea of the simple reflex was disputed: "What we have is a circuit, not an arc or broken segment of a circle. This circuit is more truly termed organic than reflex, because the motor response determines the stimulus, just as truly as sensory stimulus determines the movement (Dewey, 1896; p. 363). " PCT was developed during the 1950s by a physicist/engineer – William T. Powers who figured out how this circuit actually worked, and was first published in (Powers et al. 1960). Then it was formalized in (Powers, 1973) and revised by Powers in his latest book (2008).

PCT is a theory about how living organisms control their input instead of their output and the theory defines behavior as (merely) the control of perception through a few principles: (1) control is achieved via negative feedback; (2) control is achieved via a specific hierarchical organization of loops; (3) individuals can only control their own perception; (4) controlling others leads to conflict (5) conflict between high level control systems accounts for 'dysfunction'; (6) reorganization re-establishes control via a specific learning mechanism.

Can a PCT model account for emotions? Powers (2007) in his paper *On Emotions and PCT: A brief overview* explains how emotions can be generated through a PCT model: (1) emotion is a product of brain activity, as the brain adjusts the neurochemical reference signals that are sent from the hypothalamus into all the major organ systems via the pituitary; (2) emotion is a direct response to the disturbance, as disturbances that call control systems into action result in perceivable changes of physiological state, and those changes can be the first that one's conscious awareness knows of the presence of a disturbance; (3) In closed-loop terms, a collection of inputs, perceptions called "feelings" cause an experienced emotion, and at the same time, an output-caused change in physiological state: heart rate, respiration rate, vasoconstriction, metabolism, and motor preparedness—the "general adaptation syndrome" in the case of avoidance or attack behavior; (4) it is not the external causes but a reference signal in some high-level system which specifies a high or low intended amount of some perception that cause an emotion to happen; (5) in a high-level control system, if the current state of the perception matches what the reference signal specifies, there is a zero error signal, otherwise the mismatch leads to a nonzero error signal, resulting in a call for action to correct an error which cause an emotion; (6) a change of reference signal or a change of a disturbance can initiate a change , immediately resulting in an error signal, which gives rise both to emotional behavior and to emotional thinking; (7) the most intense negative emotions arise in connection with the largest errors and errors that humans consider the most important to correct, and their greatest intensity and duration occur when something internal or external prevents us from acting to correct the error; (8) humans use emotion-words when the degree of error is significant, important to them, resulting in being conscious of the cause while small degree of error means no use of emotion-words resulting in no identification of the cause.

In practice, PCT can inspire researchers to generate robotic emotional behavior. For instance, Moore (2007) developed a PCT based model called PRESENCE “PREdictive SENsorimtor Control and Emulation” for speech based human-machine interaction. To demonstrate PRESENCE, Moore (2007) built a Lego NXT computer model (Alpha Rex) to maximize synchrony of its own behavior with an external source. The robot can generate own rhythmic behavior, can sense its own sounds and can sense external sounds (i.e., external source). Moore’s work showed that PCT is not only a theoretical tool to explain emotional behavior but also a practical tool to predict emotional behavior.

To sum up, PCT on emotions defines emotions as one aspect of the whole integrated hierarchy of control. It admits the existence of “appraisals” (e.g., evaluation of the significance of the error signals), “adaptation” (e.g., the “general adaptation syndrome” in the case of avoidance or attack behavior), and “embodiment” (e.g., emotion is a product of brain activity), which make it compatible with other theories such as cognitive-appraisal theories, evolutionary theories, and theory of embodying emotion respectively to some extent. However, as reviewed by Tim Carey in his unpublished paper “PCT: A right turn for researchers”, various criticisms have been directed at PCT since its inception, such as it is far too ambitious to construct a theory that purports to address all human behavior, and PCT’s lack of support for a particular controlled variable.

2.2.5 Rolls’s Theory of Emotion

Rolls (1999; 2005) viewed emotions as states elicited by rewards and punishers which have particular functions. He defined a reward as anything for which an animal (which includes humans) will work, and a punisher as anything that an animal will escape from or avoid: (1) happiness is produced by being given a reward, such as a pleasant touch, praise, or winning a large sum of money; (2) fear is produced by the sound of a rapidly approaching bus, or the sight of an angry expression on someone's face; (3) frustration, anger, or sadness are produced by the omission of an expected reward such as a prize, or the termination of a reward such as the death of a loved one; (4) relief is produced by the omission or termination of a punishing stimulus such as the removal of a painful stimulus, or sailing out of danger. These examples indicate how emotions can be produced by the delivery, omission, or termination of rewarding or punishing stimuli, and go some way to indicate how different emotions could be produced and classified in terms of the rewards and punishments received, omitted, or terminated (see Fig. 2.2).

The proposal that emotions can be usefully seen as states produced by instrumental reinforcing stimuli follows earlier work by Millenson (1967),Weiskrantz (1968), Gray (1975, 1987), and Rolls (1986a, 1986b, 1990, 1999, 2000, 2005). In their proposals, instrumental reinforcers are stimuli which, if their occurrence, termination, or omission is made contingent upon the making of a response, alter the probability of the future emission of that response. Some stimuli are unlearned primary reinforcers (e.g. the taste of food if the animal is hungry, or pain); while others may become secondary reinforcers which reinforced by learning, called 'stimulus–reinforcement association'. If a reinforcer increases the probability of the emission of a response on which it is contingent, it is said to be a 'positive reinforcer' or 'reward'; if it decreases the probability of such a response it is a 'negative reinforcer' or 'punisher'. For example, fear is an emotional state which might be produced by a sound (the conditioned stimulus) that has previously been associated with an electric shock (the primary reinforcer).

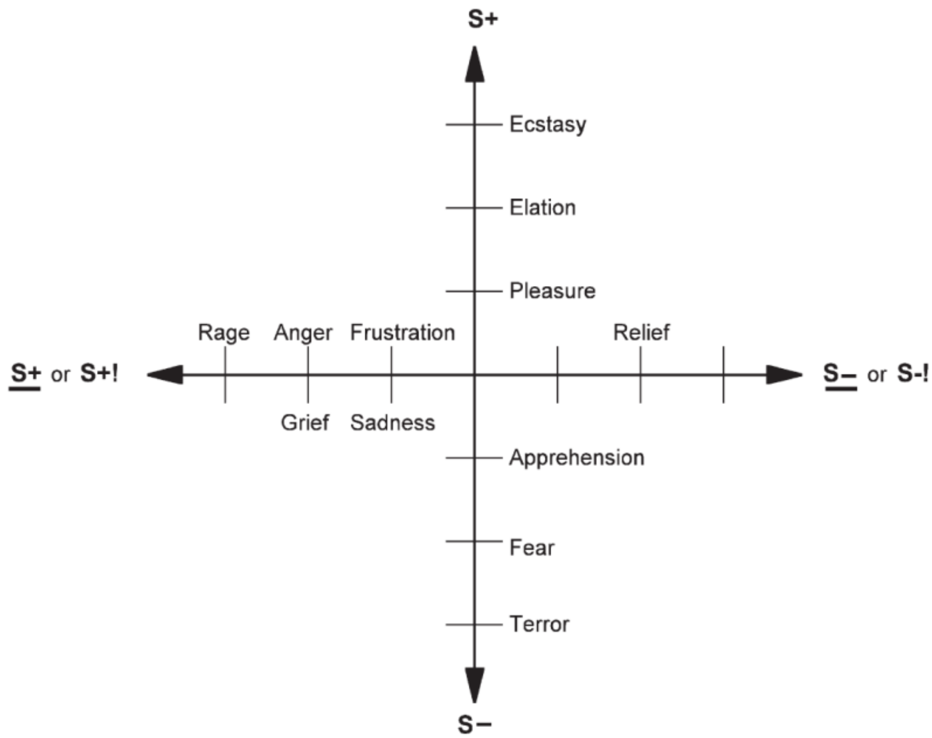


Fig. 2.2: Some of the emotions associated with different reinforcement contingencies are indicated. Intensity increases away from the center of the diagram, on a continuous scale. The classification scheme created by the different reinforcement contingencies consists with respect to the action of (1) the delivery presentation of a reward (S+), (2) the presentation of a punisher (S–), (3) the omission of a reward (S+) (extinction) or the termination of a reward (S+!) (time out), and (4) the omission of a punisher (S–) (avoidance) or the termination of a punisher (S–!) (escape) (from (Rolls, 2007), permission to reprint granted.

Can Rolls's theory of emotion account for a wide range of emotions? Rolls (2005) elaborately described how to combine the following six factors to create different emotions:

1. The *reinforcement contingency* (e.g. whether reward or punishment is given, or withheld) (see Fig. 2.2).
2. The *intensity* of the reinforcer (see Fig. 2.2).
3. Any environmental stimulus might have a *number of different reinforcement associations*. (For example, a stimulus might be associated both with the presentation of a reward and of a punisher, allowing states such as conflict and guilt to arise.)

4. Emotions elicited by stimuli associated with *different primary reinforcers* will be different.

5. Emotions elicited by *different secondary reinforcing stimuli* will be different from each other (even if the primary reinforcer is similar).

6. The emotion elicited can depend on whether an *active or passive behavioral response* is possible. (For example, if an active behavioral response can occur to the omission of a positive reinforcer, then anger might be produced, but if only passive behavior is possible, then sadness, depression or grief might occur.)

Furthermore, emotions as states elicited by rewards and punishers have functions involving particular types of neural processing and brain systems (Rolls, 2005), as summarized as follows (Rolls, 2007):

1. The *elicitation of autonomic responses* (e.g., a change in heart rate) *and endocrine responses* (e.g., the release of adrenaline).

2. *Flexibility of behavioral responses to reinforcing stimuli*. This is a crucial function of emotion in my evolutionary theory of why emotion is so important.

3. Emotion is *motivating*. For example, fear learned by stimulus reinforcement association provides the motivation for actions performed to avoid noxious stimuli. Genes that specify goals for action, for example, rewards, must as an intrinsic property make the animal motivated to obtain the reward, otherwise it would not be a reward. Thus no separate explanation of motivation is required.

4. *Communication*. Monkeys, for example, may communicate their emotional state to others by making an open-mouth threat to indicate the extent to which they are willing to compete for resources, and this may influence the behavior of other animals.

5. *Social bonding*. Examples of this are the emotions associated with the attachment of the parents to their young, and the attachment of the young to their parents. The attachment of the parents to each other is also beneficial in

species such as many birds, and humans, in which the offspring are more likely to survive if both parents are involved in the care (Rolls, 2005).

6. *The current mood state can affect the cognitive evaluation of events or memories* (Oatley & Jenkins, 1996). This may facilitate continuity in the interpretation of the reinforcing value of events in the environment.

7. Emotion may facilitate the *storage of memories*. One way this occurs is that episodic memory (i.e., one's memory of particular episodes) is facilitated by emotional states.

8. Another function of emotion is that by enduring for minutes or longer after a reinforcing stimulus has occurred, it may help to produce *persistent and continuing motivation and direction of behavior* to help achieve a goal or goals.

9. Emotion may trigger the *recall of memories* stored in neocortical representations. Amygdala back-projections to the cortex could perform this for emotion in a way analogous to that in which the hippocampus could implement the retrieval in the neocortex of recent (episodic) memories (Rolls & Stringer, 2001; Rolls & Treves, 1998). This is one way in which the memories recalled can be biased by mood states.

As can be seen, the Darwinian account of natural selection had a significant impact on Rolls's theory of emotion accounting for functions. Rolls (2005) argued that a role of natural selection is to guide animals to build sensory systems that will respond to dimensions of stimuli in the natural environment through which actions can lead to better ability to pass genes on to the next generation to increase fitness. He (2005) believed that brain design in terms of reward–punishment systems is built by genes that gain their adaptive value by being tuned to a goal for action offers.

To sum up, Rolls's theory of emotion is partially consistent with cognitive-appraisal theories (e.g., the concept of appraisal presumably involves assessing whether something is rewarding or punishing whereas the description in terms of reward or punishment adopted in Rolls's theory seems more tightly and operationally specified). As has been described above, Rolls's theory of emotion also adapted insights of evolutionary theories such that the mechanism of reward–punishment systems is the adaptation of natural selection. However, it is not consistent with social constructionist theories that argue that emotions are also cultural adaptations since it is hard to use the mechanism of reward–punishment systems to explain this perspective. Therefore, Rolls's theory of emotion focuses more on the generation of an emotion.

2.3 Summary and Conclusion

This chapter reviews the three classical theories of emotion (i.e., Evolutionary Theories, Cognitive-appraisal Theories, and Social Constructionist Theories), and other novel approaches to emotion (i.e., Theory of Embodying Emotion, Embodied Appraisal Theory, Social Functionalist Approach, Perceptual Control Theory on Emotions, and Rolls's Theory of Emotion). They are different theories focusing on different aspects of emotion: (1) what is an emotion? (e.g., Evolutionary Theories and Social Constructionist Theories mainly focus on this respect); (2) how is an emotion generated? (e.g., Cognitive-appraisal Theories, Theory of Embodying Emotion, Embodied Appraisal Theory, Perceptual Control Theory on Emotions, and Rolls's Theory of Emotion mainly focus on this aspect); (3) why do human or other species have emotions? (e.g., Social Functionalist Approach mainly focuses on this aspect, Evolutionary Theories, Social Constructionist Theories and Rolls's Theory of Emotion only partially focus on this aspect). However, the three classical theories of emotion are responsible for the development of other novel approaches to emotion, though not totally (see Fig. 2.3 below):

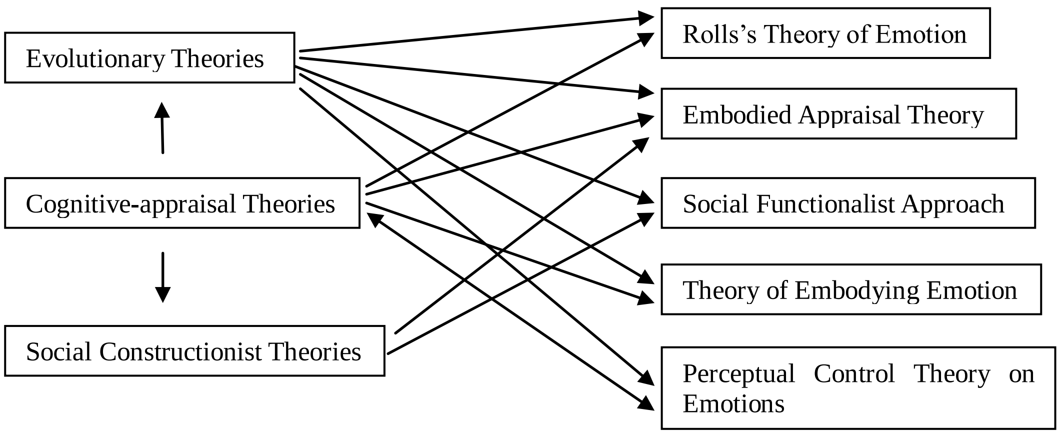


Fig. 2.3: The connections between the three classical theories and other novel approaches to emotion (the arrow indicates the direction of inspiration, for instance, Rolls's Theory of Emotion combines insights both from Evolutionary Theories and Cognitive-appraisal Theories).

From Fig. 2.3, it can be inferred that Social Constructionist Theories are not as popular as Evolutionary Theories or Cognitive-appraisal Theories, in terms of providing inspiration to other approaches. As mentioned earlier in this chapter, these three classical theories are better at giving definition of emotions than other novel approaches, as a result, they are commonly used in human or avatar emotion research. For instance, the FACS (Facial Action Coding System) (Ekman and Friesen, 1978; 2002) which is a product of Evolutionary Theories, is often used in such research in order to create appropriate facial expressions, as is Russell's circumplex model of affect (Russell, 1997; Posner et al. 2005) which is a product of Social Constructionist Theories and is frequently used in building computational models of emotion. Additionally, it is worth mentioning that Cognitive-appraisal Theories and Perceptual Control Theory on Emotions are widely used in affective computing (see historical review by Marsella, Gratch, and Petta in their paper named Computational Models of Emotion (2010)), for instance when creating synthetic emotions for agents or robots.

Other approaches to emotion such as Theory of Embodying Emotion, Embodied Appraisal Theory, Social Functionalist Approach, and Rolls's Theory of Emotion, are not widely used in affective computing, but they lend us support in understanding the generation and the purpose of an emotion more deeply. There exist differences among different theories of emotion, but it is now commonly accepted that human emotions are both biologically based and socially constructed, just as supported by the Embodied Appraisal Theory and Social Functionalist Approach. It is also supported by researchers such as Beale and Creed (2009), who argued that basic emotions like Ekman's Big Six are innate and universal, while more cognitive emotions such as concern, satisfaction, excitement and empathy, tend to be universal but can be culturally constructed.

It is worth mentioning that most of the theories of emotion deal with the generation or experience of emotions, rather than the recognition of emotions in others. How do the theories of emotion apply to the recognition of emotion in a facial expression? Yet some of these theories can still be useful in accounting for how humans recognize the emotional expressions of other people. For instance, Theory of Embodying Emotion provides good explanation of how humans recognize a facial expression of an emotion in another person. As mentioned earlier, it speculates that when a facial expression of an emotion in another person is observed, the emotion is firstly simulated in the observer's brain, supported by the promising theory of "mirror neuron system". Then the other person's facial expression is imitated without its output being observable in overt behaviour of the observer (Niedenthal et al. 2007). This bodily re-enactment of the experience of the other's state is often accompanied by self-reports of the associated emotional state (Damasio, 1994). Finally, the emotion can be recognized.

Moreover, as reviewed by Russell et al. (2003), Evolutionary Theories assume that the same message of an EE (Expression of Emotion) is encoded by the sender (or expresser) and decoded by the receiver (or observer), which guided much of the research on EEs: if (except in cases of deliberate, socially induced deception) EEs broadcast veridical information that the receiver recognizes, then researchers can focus on either the encoding or decoding side. This crucial assumption gave birth to the FACS, which relied on decoding (judgments by observers) to establish a small number of basic emotions which are expressed by a specific facial pattern. For example, Ekman & Friesen (1978,

Table 11.1) specified 65 different facial patterns they consider to be signals for anger. As can be seen, the reason humans can use the FACS to describe/measure the emotional expressions of other people, is because it is assumed that the FACS captures aspects of human knowledge (e.g., human observers can decode the facial expressions on other people's faces) about facial expressions and it becomes a comprehensive and anatomically based tool for describing/measuring any of the 44 human facial movements.

However, Social Constructionist Theories (see Russell et al. 2003) argued that evidence of the universality of the receiver's response to an EE cannot be taken as evidence of the universality of the sender's production of an EE. On the sender's side, it is not always in the sender's interest to provide veridical information. And on the receiver's side, it is unlikely that the receiver simply decodes an emotional message in any simple, reflex-like manner. As mentioned earlier, Russell et al. (2003) argued that attributing a specific emotion to the sender becomes more complex, as even when the stimuli are facial expressions of basic emotions, the context which they occur (Carroll and Russell, 1996), the receiver's current affective state (Niedenthal et al. 2000), and the gender of the sender (Widen and Russell, 2002), can color the attribution.

In next chapter, a review will be presented of how different theories of emotion in psychology are adapted to form the basis of some methods (e.g., the FACS and Russell's circumplex model of affect) to create convincing and believable synthetic emotions (e.g., Ekman's Big Six) for some emotional robots.

Chapter 3: Synthetic Robot Facial Expressions

Chapter 2 serves the purpose of giving a brief introduction of the psychology of emotion, e.g., what is an emotion, how is an emotion generated, and what is an emotion for. This chapter will review how the synthetic facial expressions of some socially interactive robots have been created based on researchers' understanding of the psychology of emotion. Insights from the psychology of emotion are used to create convincing and believable synthetic facial expressions, and even to build complex computational models of emotion for some socially interactive robots. But to create facial expressions for socially interactive robots, a computational model of emotion is not always necessary. As has been described in chapter 1, a computational model of emotion is useful in arbitrating a robot's internal affect states and in coordinating a robot to respond to the external environment appropriately. More often, tools like the FACS and techniques like mapping facial expressions from videos of people to a robot head, can be used to create facial expressions for lots of socially interactive robots. Examples of these robots will be given in this chapter.

This chapter focuses on how to create convincing and believable synthetic facial expressions of socially interactive robots.

Why does this thesis focus only on the faces of robots? It does so for two reasons. First of all, in everyone's daily life, the face plays a very important role in the expression of character, emotion and/or identity (Cole, 1998). Actually, a study by Mehrabian (1968) showed that facial expression is a major modality in human face-to-face communication: only 7% of affective information is transferred by spoken language, that 38% is transferred by paralanguage and 55% of transfer is due to facial expressions. Secondly, facial expressions of robots play an important role in human-robot interaction. Actually, in the field of human computer interaction (HCI), research by Reeves and Nass (1996) has demonstrated that humans (whether computer experts, lay people, or computer critics) generally treat computers as they might treat other people. Their numerous studies indicate that a social interface may be a truly universal interface (Reeves and Nass, 1996). Breazeal (2003a) argued that humanoid robots (and animated software agents) are arguably well suited to this: sharing a similar morphology, for instance facial expression, they can communicate in a manner that supports the natural communication modalities of humans.

Why are socially interactive robots focused on here?

Before knowing the importance of socially interactive robots, it would be more understandable if one knew the definition of socially interactive robots. In a paper called *A Survey of Socially Interactive Robots: Concepts, Design, and Applications*, Fong et al. (2002) used the term “socially interactive robots” to describe robots for which social interaction plays a key role. They drew the “human social” characteristics of socially interactive robots as follows: (1) express and/or perceive emotions; (2) communicate with high-level dialogue; (3) learn/recognize models of other agents; (4) establish/maintain social relationships; (4) use natural cues (gaze, gestures, etc.); (6) exhibit distinctive personality and character ; (7) may learn/develop social competencies.

Fong et al. (2002) also stressed the importance of socially interactive robots: socially interactive robots are important for domains in which robots must exhibit peer-to-peer interaction skills, either because such skills are required for solving specific tasks, or because the primary function of the robot is to interact socially with people. The authors demonstrated two ways of building socially interactive robots: a “biologically inspired approach” in which designers try to create robots that internally simulate, or mimic the social intelligence found in living creatures and a “functionally designed approach” in which the goal is to construct a robot that outwardly appears to be socially intelligent, even if the internal design does not have a basis in science.

Moreover, socially interactive robots are employed in many areas nowadays, Fong et al. (2002) listed the following examples: they can be used as test subjects, for instance, robots have been used to examine, validate and refine theories of social and biological development, psychology, neurobiology, emotional and non-verbal communication, and social interaction, they can have great functional value when they provide concrete services for humans, they can be treated as animatronic children's dolls, mobile social companions - quadruped and wheeled personal robots, or interactive goal-directed tools, and moreover, they can be used in rehabilitation and therapy. In a word, socially interactive robots to humans, are partners that humans can interact and cooperate with, rather than just tools.

3.1 Social Robot Embodiment

In general, there are human-machine interaction systems that can be either embodied (the human interacts with a robot or an animated avatar) or disembodied (the human interacts through speech or text entered at a keyboard). Breazeal (2003a) believed that embodied systems have the advantage of sending para-linguistic communication signals to a person, such as gesture, facial expression, intonation, gaze direction, or body posture. As a result, these embodied systems must address the issue of sensing the human, often focusing on perceiving the human's embodied social cues. According to Dautenhahn et al. (2002), the embodiment of a robot can be defined as "that which establishes a basis for structural coupling by creating the potential for mutual perturbation between system and environment". And the more a robot can perturb an environment, and be perturbed by it, the more it is embodied, which also means that social robots do not necessarily need a physical body. In a review chapter titled 'S*ocial robots that interact with people'* by Breazeal et al. (2008), a unifying characteristic is defined as that social robots engage people in an interpersonal manner, communicating and coordinating their behavior with humans through verbal, nonverbal, or affective modalities. Nowadays, social robots have a rich variety of embodiments, varying from humanoid, android, creature-like, or simply non-humanoid and non-zoomorphic in form.

3.1.1 Humanoids

A number of socially interactive humanoid robots have been equipped with arms and hands that are designed to exhibit human-like gestures such as pointing, shrugging shoulders, shaking hands, or giving a hug, or with mechanical faces to communicate with humans via facial expressions. As a result, these humanoids can participate in whole body social interaction with people such as dancing, walking hand-in-hand, playing a musical duet, or transferring skills to unskilled persons (Breazeal et al. 2008).

Examples of socially interactive humanoid robots include: (a) humanoid robots developed at Waseda University: a flautist robot WF-4RII (Solis et al. 2006), WABIAN-2 (Ogura et al. 2006), and WE-4RII (Miwa et al. 2004); (b) Robovie III developed by the Advanced Telecommunications Research Institute International (ATR) Intelligent Robotics and Communication Laboratories, which is able to gesture with its arms and give a hug (Roccella et al. 2004; Miwa et al. 2004); (c) quest for curiosity (QRIO), a small biped entertainment robot previously developed by Sony, which is very well known for its impressive dancing ability (Tanaka et al. 2006); (d) a Fujitsu HOAP-2 humanoid robot that can recognize and reproduce social cues, that was taught by Sylvain Calinon and Aude Billard (2006) to use gaze directions as turn-taking cues, turn-yielding cues, and turn-requesting cues.

3.1.2 Androids

Whereas many humanoid robots have a mechanical appearance, android robots are designed to have a very human-like appearance with skin, teeth, hair, and clothes. As Breazeal et al. (2008) stated, a design challenge of android robots is to avoid the uncanny valley where the appearance and movement of the robot resemble more of an animate corpse than a living human. Designs that fall within the uncanny valley elicit a strong negative reaction from people (Mori, 1970).

Examples of androids include: (a) one of the earliest face robots developed at the Science University of Tokyo (Iida et al. 1998); (b) Geminoid HI-1 developed by the ATR Intelligent Robotics and Communication Laboratories (Shimada et al. 2006); (c) ROMAN, developed at the University of Kaiserslautern (Berns and Hirth, 2006); (d) the Waseda DOCOMOFACE Robot No. 2 developed by Waseda University (Hayashi et al. 2006); (e) Nexi (Alonso, 2009), developed by MIT’s Personal Robot Group, is a complete mobile manipulator robot augmented with rich expressive abilities; (f) Kaspar (Hatice Kose-Bagci et al. 2008), a therapeutic robot that demonstrated an emergent ability to develop the dynamics of turn-taking interaction in drumming games; (g) a realistic robot head called “Jules” (Jaeckel et al. 2008a), developed by *Bristol Robotics Laboratory.*

3.1.3 Creature-like Social Robots

Creature-like social robots are aesthetically and behaviorally inspired by animals: they can entertain people with touch-based communication ability, they can be therapeutic companions with a more organic appearance, or they can have a more fanciful appearance, melding anthropomorphic with animal-like qualities.

Examples of creature-like social robots include: (a) AIBO, the robotic dog previously developed by Sony (Fujita, 2004); (b) Paro, the therapeutic seal robot developed at AIST (Wada et al. 2005); (c) Mel, the conversational robotic penguin developed at MERL (Sidner et al. 2005); (d) Leonardo, developed at the MIT Media Lab (Breazeal et al. 2004); (e) iCat (van Breemen et al. 2005), is a robot companion cat produced by the Phillips Research laboratories; (f) EmI (Saint-Aimé et al. 2009), is a teddy bear-like therapeutic companion robot, which is currently the 2nd generation of a design developed by Valoria Laboratory - University of Bretagne Sud, France; (g) Probo, the appearance of which represents an imaginary animal based on the ancient mammoths, is a research platform to study human-robot interaction with a focus on non-verbal communication, the design of which is adapted to the needs of hospitalized children (Saldien et al. 2006).

3.1.4 Non-humanoid and Non-zoomorphic Social Robots

Many social robots are not overtly humanoid or zoomorphic, but still capture key social attributes serving for particular social purposes: (1) Kismet (Breazeal, 2002), one of the best-known and pioneering social robots developed at the MIT Artificial Intelligence Lab, was built for establishing interaction between Kismet and a human by using the parent-child relationship during early communication as inspiration; (2) Emo Shun, an emotional robot head having a very expressive mechanical face with anthropomorphic features like large blue eyes, built by Noel Sharkey from the University of Sheffield, was on display at Birmingham Millennium Thinktank Science Museum to capture the audience reaction; (3) Keepon (Kozima, 2006), a dancing robot having a simplistic face that uses a classic animation technique called squash and stretch for expression of the body, developed by National Institute of Information and Communications Technology (NiCT) (Japan); (4) Autom (Kidd and Breazeal, 2008), designed to play a role as a weight loss coach to guide humans to loss their weight, showed that developing long-term close relationship between human and robot was practicable; (5) mobile social robots like Valerie with a graphical face on a liquid crystal display (LCD) screen (Gockley et al. 2006), developed at Carnegie Mellon University, which was built to be a robotic receptionist; (6) the SmartKom system (Alexandersson et al. 2004), built in Germany, provided full symmetric multimodality in a mixed-initiative dialogue system with an embodied conversational agent, that can understand and represent not only the user's multimodal input, but also its own multimodal output; (7) some social robots have no overt social features like faces or eyes, but rely purely on language-based communication.

3.2 Creating Recognizable Facial Expressions for Social Robots

As described in Chapter 2, the FACS (Facial Action Coding System) (Ekman and Friesen, 1978; 2002), is a comprehensive and anatomically based tool for describing/measuring any human facial movement. But it has been found to be a reliable tool to create recognizable facial expressions for some social robots.

Before applying the FACS to any robot face, it is important to know that the facial appearance of a robot makes a difference in people's perception of its facial expressions. First of all, facial morphology is important for the interpretation of and reactions to emotional facial expressions of human beings, because humans not only have expectations of an expresser's probable emotional reactions after receiving information from the face but also use some of the features to recognize certain emotional expressions and derive personality (Ursula et al. 2009). Secondly, the "Media Equation" states that people not only can but do treat computers, televisions, and new media as real people and places (Reeves and Nass, 1996). Bartneck et al. (2005) extended the idea of "Media" to robots, since they often have an anthropomorphic embodiment and human-like behavior. Finally, people can perceive the signaled emotions of robots via facial expressions (Breazeal, 2003a), and Saldien et al. (2010) drew a conclusion that the facial features serving as social cues for human beings to recognize the synthetic robot emotions on a robot face played an important role in perception of the expressions of emotions.

Despite the large variation in the appearance of robot faces, the FACS contributes significantly to the generation of convincing facial expressions of some emotional robots. For instance, the FACS was widely adapted to create at least six basic and discrete facial expressions (joy, sadness, anger, fear, disgust and surprise) for some non humanoid robotssuch as Feelix (Canamero and Fredslund, 2000), Kismet (Breazeal, 2002), Aryan (Mobahi et al. 2003), EDDIE (Sosnowski et al. 2006), and Probo (Saldien et al. 2010). The FACS was initially developed as a set of guidelines for recognizing the facial expressions of humans but has been found to be a reliable tool for creating believable versions of these basic facial expressions (e.g., overall recognition rate of the basic facial expressions for Probo was 84%, followed by Kismet with 73%, while Eddie achieved a bit lower rate with 57%.), even though some of these evaluations were based on posed but not spontaneous robot facial expressions and some facial expressions (fear, and happy+surprised) were harder to make convincing (Breazeal, 2002; Saldien et al. 2010).

Moreover, to generate dramatic but smooth shifts from one emotional state of a robot to another, Russell's circumplex model of affect containing two dimensions: valence and arousal (Russell, 1997) was adapted to construct an emotion space for Kismet and its new version (Posner et al. 2005) was used for EDDIE and Probo. Although it is not necessary to use a circumplex model of affect to create facial expressions of a robot, such a model is useful for improving the life-likeness of a robot. This is because not only the facial expression itself but also how this expression changes over time contains information as dramatic but smooth shifts from one emotional state to another can make robot expressions become more life-like (Breazeal, 2003a).

In summary, what these emotional robots look like and how the facial expressions are shown on their faces, can determine how humans perceive their facial expressions. The following examples, including two non android robots Kismet and Probo and two androids Geminoid F and Jules, incorporate a more detailed account of how the robot faces were equipped with rich features and how the facial expressions of these robots were created.

3.2.1 Examples of Non-Android Robots

A. Kismet

Kismet (Breazeal & Scassellati, 2000) is an expressive robot developed by MIT that has 4 perceptual modalities (speech, body posture, facial display, and gaze control). The main purpose of the Kismet project was to explore the origins of social interaction and communication in people, namely that which occurs between caregiver and infant, through extensive computational modeling guided by insights from psychology and ethology (Breazeal et al. 2008).

An overview of the project can be found in (Breazeal, 2002). The robot is about 1.5 times the size of an adult human head and has a total of 21 Degrees of Freedom (DoF). Three DoF direct the robot's gaze, another three control the orientation of its head, and the remaining 15 move its facial features (e.g., eyelids, eyebrows, lips, and ears). Kismet is equipped with a total of four color CCD cameras (there is one narrow field of view camera behind each pupil and the remaining two wide field of view cameras are mounted between the robot's eyes), so that it can visually perceive the person who interacts with it. In addition, Kismet has two small microphones (one mounted on each ear). A lavalier microphone worn by the person is used to process their vocalizations. These motor systems let Kismet automatically adjust its visual and auditory detectors toward the stimulus source.

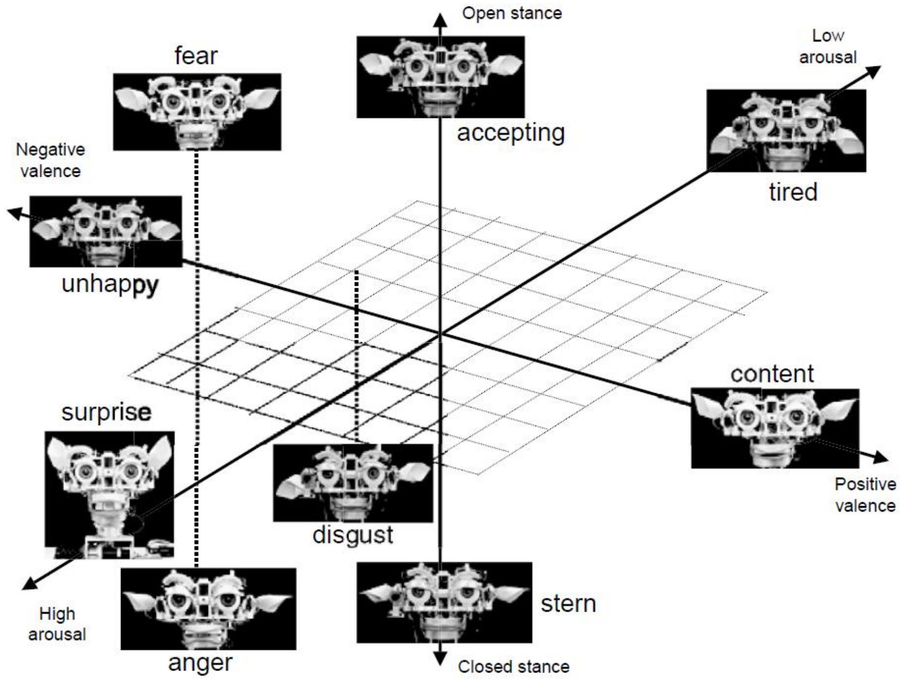


Fig. 3.1: This diagram illustrates where the basis postures are located in affect space. Permission from Emotion and Sociable Humanoid Robots (Breazeal, 2003) to reprint granted.

Kismet is capable of generating a continuous range of expressions of various intensities by blending the basis facial postures, owing to the use of an interpolation-based technique over a three dimensional, componential affect space (arousal, valence, and stance) (Breazeal, 2002, see Fig. 3.1). This computational model of emotion was tested with 5 primary emotions (anger, disgust, fear, sadness, happiness) and three additional ones (surprise, interest, and excitement). It was inspired by Russell's circumplex model of affect containing two dimensions: valence and arousal (Russell, 1997). Moreover, to generate dramatic but smooth shifts from one emotional state of a robot to another, a three dimensional, componential affect space adapted from Russell's circumplex model of affect (Russell, 1997), was constructed for Kismet. This three dimensional affect space, according to Breazeal (2002), resonates well with the work of Smith and Scott (1997). They posit a three dimensional space of pleasure-displeasure (maps to valence here), attentional activity (maps to arousal here), and personal agency/control (roughly maps to stance here). They also posit a fourth dimension that relates to the intensity of the expression. Although Kismet's dimensions do not map exactly to those hypothesized by

Smith and Scott, Kismet's model conceptualizes the idea of combining meaningful facial movements in a principled manner to span the space of facial expressions, and to also relate them in a consistent way to emotion categories. According to Breazeal (2002), several advantages can be obtained from this affect space, such as promoting smooth trajectories through affect space, allowing the robot's facial expression to reflect the nuance of the underlying assessment, and lending clarity to the facial expression since the robot can only be in a single affective state at a time (by choice) and hence can only express a single state at a time.

To design the basis facial postures of Kismet, the componential model of facial expressions theorized by Smith and Scott (1997) was adapted, whereby individual facial features move to convey affective information. The raised brows, for instance, convey attentional activity for both fear and surprise. They proposed a possible mapping of facial movements to affective dimensions. Using the Facial Action Coding System (FACS) developed by Ekman and Friesen (1978), Smith and Scott (1997) have compiled mappings of FACS Action Units (AU) to the expressions corresponding to anger, fear, happiness, surprise, disgust, and sadness. AU expresses a motion of minimal muscle to realize one of the 44 kinds of basic operation, with 14 AUs used to express the emotions of anger, disgust, fear, joy, sadness, and surprise.

To explore questions such as *How do people identify Kismet's facial expressions with human expressions?* and *Do they map Kismet's distinctive facial movements to the corresponding human counterparts?* Breazeal (2002) conducted an experiment, in which she asked naive subjects to perform a comparison task where they compared color images of Kismet's expressions with a series of line drawings of humans from (Faigin, 1990). The result of the experiment indicated that the subjects seem to intuitively match Kismet's facial features to those of the line drawings, and interpreted their shape in a similar manner.

However, the line drawing study did not ask the subjects what they thought the robot was expressing. To explore this issue, a color-image-based evaluation was continuously conducted by Breazeal (2002). In the experiment, the subjects chose from ten possible labels (accepting, anger, bored, disgust, fear, joy, interest, sorrow, stern, surprise) that best matched the picture of Kismet displaying one of seven expressions (anger, disgust, fear, happiness, sorrow, surprise, and a stern expression) in a forced choice questionnaire. The subjects' responses were significantly above random choice, ranging from 47% to 83%, for example, joy, sorrow, and surprise were scored with over 80% whereas fear was only recognised with 47% accuracy.

In Breazeal 's view (Breazeal, 2002), the still image and line drawing studies were useful in understanding how people read Kismet's facial expressions, but they reveal little about expressive posturing, as humans and animals do not only express with their face, but with their entire body. She explored this issue with a video evaluation in which the subject was asked to select a word that best described the robot's expression (anger, disgust, fear, joy, interest, sorrow, or surprise) from one of the seven videos in all (anger, disgust, fear, joy, interest, sorrow, and surprise) in a forced-choice paradigm. The subjects performed significantly above chance, with overall stronger recognition performance than on the still images alone: the video segments for the expressions of anger, disgust, fear, and sorrow were correctly classified with a higher percentage than the still images.

To sum up, even though Kismet lacks many of the facial features of humans (most notably, skin, teeth, and nose), the Kismet project showed that Kismet's expressions are readable by people with minimal to no prior familiarity with the robot. The Kismet project also emphasized the importance of constructing an emotional space allowing a smooth shift between discrete emotions, although there was no comparison of the believability of the expressions with and without the smooth shift. Moreover, the Kismet project demonstrated that the robot could not only arbitrate its internal affect states but also interact socially with humans by using its computational model of emotion (see Breazeal, 1998; Breazeal and Aryananda, 2002 to know how the robot expresses its "emotive" state, and how it regulates its interaction with people while arbitrating its own emotional states). The model itself is not reported here because my focus is more on how to create facial expressions of a robot than on questions about how to create a robot that responds socially to a human.

B. Probo

Probo is a research platform for the study of human-robot interaction with a focus on non-verbal communication, the design of which is adapted to the needs of hospitalized children (Saldien et al. 2006). The appearance of the robot represents an imaginary animal based on the ancient mammoths, which has less resemblance to existing animals or humans aiming to avoid Mori's problem of the "uncanny valley" (Mori, 1981), according to (Saldien et al. 2010).

The first prototype of the robot had 20 Degrees of Freedom (DOF) to obtain a fully-actuated head and trunk. By moving its head (3 DOF), eyes (3 DOF), eyelids (2 DOF), eyebrows (4 DOF), ears (2 DOF), trunk (3 DOF) and mouth (3 DOF) the robot is able to express its emotions (Saldien et al. 2006). The robot is equipped with a range of sensory input devices, such as a digital camera, microphones, position sensors, temperature sensor, touch sensors under the fur, giving the robot the ability to capture the stimuli from its environment (Saldien et al. 2006). More information on the internal mechanics of Probo is described in earlier work of Saldien et al. (2008).

For the display of the emotions, most of the DOF in the face are based on the Action Units (AU) from the FACS (Ekman and Friesen, 1978). Because Probo (Saldien et al. 2008) does not have a human face and in order to simplify the design, some of the AUs are missing, others are replaced and some are added. The lack of the lower eyelid and a fixed upper lip lead to missing AUs: the AUs regarding the nose movements (AU 9 and 11) were replaced by the movement of the 3 DOF trunk. The movement of the ears and the greater visual influence of the trunk added extra gestures to express the emotions.

Similar to the Kismet project that used a three dimensional, componential affect space (arousal, valence, and stance) (Breazeal, 2002), the Probo project also has a computational model of emotion. Only two dimensions: valence and arousal are used to construct an emotion space in the Probo model (Saldien et al. 2008), based on the circumplex model of affect defined by Russell (Posner et al. 2005).

Like the Kismet project, the Probo project also involved a few recognition tests to find out whether people who interact with the robot recognize the correct emotion behind its facial expressions. Some of the tests are heavily impacted by the Kismet project. Details of all the 7 tests are elaboratively described in (Saldien et al. 2010). In brief, during the design phase, a virtual model of Probo was built to simulate the facial expressions used in the first five recognition tests (T1-T5). After the prototype was built, additional tests were performed with pictures of the physical robot (T6 and T7, see Fig. 3.2). A closed answer format was used in similar tests performed with other robot faces (Breazeal, 2002; Canamero and Fredslund, 2000; Mobahi, 2003; Sosnowski et al. 2006). According to (Saldien et al. 2010), they used the same format inT5, T6 and T7 for testing the robot Probo, for a good comparison.

Fig. 3.2: The 6 basic facial expressions of the covered prototype used in T7 (happy, surprise, sad, anger, fear and disgust). Permission from (Saldien et al. 2010) to reprint granted.

In more detail, a survey performed by Breazeal (2002) evaluating the expressive behavior of Kismet was adapted for Probo's first pilot study with children (T1) and adults (T2). In T1 and T2, the subjects were asked to perform a comparison task where they compared 8 color images of the virtual model with a series of line drawings of human expressions from (Faigin, 1990). In the second study, people were asked to match the 8 color images with 8 given answers (words) identifying each emotion by using a multiple-choice-test. For this study a paper version was used to test the children (T3) and an electronic version aiming to test more people was used to test the adults (T4). To facilitate comparisons between the expressions of Probo with other similar robot projects, only 6 basic facial expressions (happy, surprise, sad, anger, fear and disgust) were included in the last 3 tests. These tests include the 3 different display modi as there are; the virtual model, the uncovered prototype and the covered prototype. In these tests the motion of the trunk was used to emphasize the expressions, based on the AU of the human nostrils. Evaluations for the projects EDDIE (Sosnowski et al. 2006), Kismet (Breazeal, 2002), Aryan (Mobahi, 2003) and Feelix (Canamero and Fredslund, 2000) were compared with the covered prototype of Probo in T7.

Saldien et al. (2010) drew the following conclusions on the basis of the findings in all 7 experiments : (1) from the first tests (T1-T4), it is better to use a list of words rather than a list of drawings to measure the recognition of emotions in facial expressions; (2) as for T5, an overall recognition rate of 88% was achieved for the virtual model after incorporating the trunk movement in the facial expressions; (3) The results of T6 and T7 showed that the recognition of the emotional states of Probo was substantially higher than the others (e.g., overall recognition rate of the basic facial expressions for Probo was 84%, followed by Kismet with 73%, while Eddie achieved a slightly lower rate with 57%.); (4) in all the projects the recognition of fear has the lowest score.

Saldien et al. (2010) provided further discussion of (3) and (4) above. In terms of (3), they argue that it is because Probo has more facial features than other robots, and facial features, serving as social cues for human beings to recognize the synthetic robot facial expressions on a robot face, play an important role in the perception of the emotional expressions. As for (4) above, it was consistent with the finding reported elsewhere that the overall recognition rate of fearful human facial expressions is lower than that of other basic emotions (Fairchild et al. 2009).

To sum up, the project Probo was strongly influenced by the project Kismet. Both projects emphasized the use of Action Units from the FACS (Ekman and Friesen, 1978) as a tool to create facial expressions of the robot and the adapting of an emotional space to allocate discrete emotions. Where the Probo project differs from the Kismet project is that currently no computational model of emotion is developed for Probo, as it focuses more on how people view its facial expressions rather than how it responds to the people who view its face.

3.2.2 Examples of Android Robots

A. Geminoid F

In 2005, Ishiguro defined *Android Science* as a cross- interdisciplinary framework. Before he proposed the framework, he argued that a humanlike appearance and humanlike behavior of the robot are important for developing and evaluating interactive robots. Hence, he suggested that there are two necessary approaches to tackle the problem of appearance and behavior: one from robotics and the other from cognitive science. The approach from robotics tries to build very humanlike robots based on knowledge from cognitive science. The approach from cognitive science uses the robot for verifying hypotheses for understanding humans. The two approaches therefore constitute his framework of android science.

Ishiguro (2005) continued to describe that there are three steps to develop androids from the robotics approach before using androids to conduct cognitive tests: (1) create very humanlike appearance for androids; (2) develop mechanisms for humanlike movements and reactions of androids; (3) equip androids with humanlike perceptual abilities such as with the help of ubiquitous/distributed sensor systems.

When an android that not only looks like a human and behaves like a human but also perceives like a human is developed, it can be used in cognitive studies. The effects of an android's anthropomorphic appearance and its body movements have been investigated in a number of empirical studies with Geminoid HI-1 robot, but none of them have focused on evaluating facial displays of emotion for the android. Geminoid HI-1, a facial expressions-capable robot, was developed so as to closely resemble the outer appearance of Ishiguro, its creator, which made it a revolutionary machine in its time. However, according to (Becker-Asano and Ishiguro, 2011), Ishiguro's new android robot “Geminoid F” has a higher degree of human-likeness, and was modeled to resemble her human counterpart’s outer appearance to the finest detail, which together with its sophisticated mechanical design, permits the creation of diverse facial expressions. Therefore, the development of “Geminoid F” has concentrated on its ability to perform sophisticated facial expressions. And its facial expressivity can be compared with that of its model person since it was also modeled after a real person.

Compared with Geminoid HI-1, the Geminoid F is capable of offering more while needing fewer resources to be built and function. The degrees of freedom, for instance, have been reduced from 46 to only 12 on this new model: seven are located in its face, three in its head and neck, and two in its upper body. However, the limbs of Geminoid F are immobile even though it is a tele-operated robot like Geminoid HI-1, which according to (Becker-Asano and Ishiguro, 2010), make it have the advantage that the controllers for the pneumatic actuators could be integrated into its body such that only one air pressure and one controller cable needs to be connected to Geminoid F.

In (Becker-Asano and Ishiguro, 2011), two online surveys designed to evaluate Geminoid F’s facial display of five basic emotions were described. Each survey had versions in English, German, and Japanese for the purpose of looking for inter-cultural differences. Six digital pictures of Geminoid F’s face featuring the basic emotions angry, fearful, happy, sad, and surprised plus a neutral expression, were presented in the first survey (see Fig. 3.3). They were realized by manually adjusting the actuators through a software interface. The second survey was very similar to the first one with the only difference being that it presented pictures of the human model used for Geminoid F instead of the robot.

Fig. 3.3: The android robot Geminoid F (top row) and its model person (bottom row) portraying five basic emotions and a neutral expression, from left to right: angry, fearful, happy, neutral, sad, surprised. Copyright © 2011 IEEE, reprinted with permission from (Becker-Asano and Ishiguro, 2011).

According to (Becker-Asano and Ishiguro, 2011), they found that: (1) the emotional expressions “fearful” and “surprised” were often confused, and also that many Japanese participants seemed to confuse “angry” with “sad” in contrast to the German and English participants; (2) although similar facial displays portrayed by the model person of Geminoid F achieved higher recognition rates than those of the android overall, the fearful expression was also often not recognised; (3) the happy, neutral and sad facial expressions Geminoid F were well recognised , but surprised and angry expressions only achieved a moderately successful recognition rate, whilst their portrayal of a fearful expression failed the test. Therefore, they concluded that improving the android’s expressiveness especially around the eyes would be a useful next step in android design. It was suggested that Geminoid F’s rather limited ability to change its face around the eyes was responsible for the more ambiguous ratings of the Asian (esp. Japanese) participants, who tend to focus more on that facial region and less on the mouth (Yuki et al. 2007). They also indicated that an evaluation of dynamic facial expressions of Geminoid F in future research could complement the results found in these two surveys.

However, it is worth mentioning that even though Geminoid F has a more human-like appearance with synthetic skin, hair, eyes and teeth than that of Kismet (Breazeal, 2002) and Probo (Saldien et al. 2010), Geminoid F did not gain a better recognition rate of basic facial expressions than Kismet and Probo.

For instance, the overall average recognition rate of the six facial expressions for Probo is 84% (no neutral expression but a disgusted facial expression scoring 87% is included), and for Kismet is 73% (no neutral expression but a disgusted facial expression scoring 71% is included), whereas for Geminoid F it was only 61% (no disgusted facial expression, but a neutral expression scoring 83.9% was included). More importantly, the overall recognition rate of each of the five basic facial expressions (angry, fearful, happy, sad, and surprised expressions) for Probo or for Kismet was higher than that of Geminoid F. It seems that a more human-like appearance did not enhance the overall recognition rate of basic facial expressions. Future research of Geminoid F could do with providing more reasons for this result beyond Geminoid F's rather limited ability to change its face around the eyes.

In addition, in (Becker-Asano and Ishiguro, 2011), little was reported about how the facial expressions of Geminoid F were created, for instance, did they implement the FACS as well? The paper only stated "They were realized by manually adjusting the actuators through a software interface". It would be better if future research of Geminoid F was to state clearly how its facial expressions are created, just as the project of Kismet (Breazeal, 2002) and the project of Probo (Saldien et al. 2010) did. This might also help the project to explain why Geminoid F did not gain high recognition rate of basic facial expressions.

B. Jules

Jules is an animatronic head produced by the US roboticist David Hanson (website: www.hansonrobotics.com) within a project, called "Human-Robot Interaction", which was developed at the Bristol Robotics Laboratory (BRL), run by the University of the West of England and the University of Bristol. Jules has 34 internal motors covered with flexible rubber ('Frubber') skin. The servo motors are controlled by servo controllers that receive desired positions for each servo motor via an RS232 interface at up to 30fps (frames per second). The robotic head can automatically mimic a person's expressions simply by watching their face through a video camera (see Fig. 3.4). The Bristol team successfully developed software to transfer expressions recorded by the video camera into commands to make those servos produce similarly realistic facial movements.

They proposed models which map from facial expressions to a robots muscle or servo space using explicit functional relationship such as a Gaussian Process Regression (Jaeckel et al. 2008a), or Linear Partial Least Squares (Jaeckel et al. 2008b), and Shared Gaussian Process Latent Variable Models for the detection and modelling of ambiguities in data sets of human and robotic facial behaviour (Henrik et al. 2009). They showed that the mapping is efficient, and robust to tracking errors.

Fig. 3.4: Well mapped mouth shapes: The top and bottom of each row show the actress's facial expression and the resulting poses of the robot head, respectively. Reprint with permission from (Jaeckel et al. 2008b).

The current stage of their project does not involve the robot in any facial recognition experiment. As described above, similar facial displays portrayed by the model person of Geminoid F achieved higher recognition rates overall than that of Geminoid F (Becker-Asano and Ishiguro, 2011). Therefore, one might expect that the recognition rates of facial expressions of Jules would be decently high, based on good quality of mapping from video footage to the robot head.

3.2.3 Summary of the Methods

Breazeal (2002) reported research using the Kismet robot that demonstrated that even although that robot has a mechanical appearance and lacks many of the facial features of humans (most notably, skin, teeth, and nose), its facial expressions are still readable by people with minimal to no prior familiarity with the robot. The three dimensional, componential affect space (arousal, valence, and stance) which has roots in Russell's circumplex model of affect containing two dimensions: valence and arousal (Russell, 1997), and Smith and Scott (1997)'s theory of mapping of Action Units (AU) of the Facial Action Coding System (FACS) (Ekman and Friesen, 1978) to the expressions corresponding to basic emotions, were well used in the Kismet project.

The author of this thesis has also looked at the Probo robot (Saldien et al. 2010), another socially interactive robot that does not look as mechanical as Kismet and has more facial features. The Probo project, following the path of the Kismet project, also created its facial expressions with the help of Action Units (AU) and a emotional space based on Russell's circumplex model of affect (Posner et al. 2005). Similarly, the Probo project (Saldien et al. 2010) successfully created convincing and believable facial expressions for the robot, and obtained an overall recognition rate of the facial expressions of which was even better than that of Kismet.

The Kismet project, the Probo project, and other social robot projects involving the FACS to create robotic facial expressions have shown that convincing and believable robotic facial expressions can be created with the help of the FACS. As mentioned in Chapter 2, the FACS rested on the assumptions drawn from Evolutionary Theories that the same message of an EE (Expression of Emotion) is encoded by the sender (or expresser) and decoded by the receiver (or observer). In other words, the reason humans can use the FACS to describe/measure the emotional expressions of other people, is because Evolutionary Theories assume that facial expressions, e. g. a smile, can be decoded by a receiver, and the FACS describes and categorizes these human facial expressions. The FACS is such a comprehensive and anatomically based tool that not only can human decoders use it to describe/measure any of the 44 human facial movements, but also can human coders use it to manually encode nearly any anatomically possible facial expression. The computer avatar research (see Creed and Beale, 2008; Noël et al. 2009; Mower et al. 2008; 2009) already showed that facial expressions of avatars coded by human programmer can be recognized by human observers as intended. If the FACS description is then used to generate expressions for a robot, those expressions are likely to be recognized as intended. And results from the recognition tests in robotic projects such as the Kismet project, have indicated that when the FACS is used to create robot facial expressions, those expressions can be recognized as intended by human observers. Consequently, a conclusion is arrived at that theories of emotion (e.g., Evolutionary Theories) as described in Chapter 2 are also relevant to the recognition of emotional expressions of robots, since the same message of a robotic EE (Expression of Emotion) can be encoded by the programmer and decoded by the receiver (human subject).

Androids, on the other hand, are more human-like in terms of their appearance and behaviour. Ishiguro’s creations of android robots have pushed the human-likeness of these robots to a very high degree, for instance, his new android robot “Geminoid F”. According to Becker-Asano and Ishiguro (2011), at first sight and from a distance it is difficult to tell Geminoid F and its human model person apart and this difficulty remains even when both of them are moving slightly. Surprisingly, although its appearance is more human-like than that of Kismet or Probo, the project failed to achieve an overall recognition rate of the facial expressions corresponding to basic emotions as high as that of the Kismet project and the Probo project. However, the project showed that similar facial displays portrayed by the model person for Geminoid F achieved higher recognition rates overall than that of Geminoid F, which indicate a good prospective for techniques that involve mapping from video footage to the robot head. For instance, the Jules project (Jaeckel et al. 2008a; 2008b; Henrik et al. 2009) demonstrated that, by using certain models which map from facial expressions to a robot's muscle or servo space, they could achieve good quality of mapping that allows the robot to automatically mimic a person's expressions simply by watching their face through a video camera.

3.3 "Social" Aspects of the Design of a Human-robot Interaction

In the review presented above, it was found that robot faces can be equipped with rich features and that socially interactive robots can be developed to create convincing robot facial expressions. But what are "social" aspects of the design of a human-robot interaction? Mutlu et al. (2006) presented the critical "social" aspects of the design of a human-robot interaction, namely robot attributes (e.g. appearance, character, styles of speech, motion), user attributes (e.g. gender, age, race, education level, income level), and task structure (e.g. collaborative, competitive, creativity, planning, intellective, mixed-motive). Since these do seem to capture the main aspects of social human robot interaction, robot attributes, user attributes, and task structure will be reviewed in this section.

3.3.1 Robot Attributes

Researchers in human-robot interaction (HRI) have mainly focused on the capabilities of sociable emotional robots themselves, for example, their capability to detect the affective states of users, their capability to use synthetic facial expressions to express recognizable emotions, speech and textual content, and in particular, their ability for imitating and social learning. MIT's Personal Robot Group has been a pioneer in developing outstanding capabilities of sociable emotional robots, such as Kismet (Breazeal, 2002), Leonardo (Brooks et al. 2004), and Nexi (Named #17 in TIME's 50 Best Inventions of 2008, see (Alonso, 2009), all of which have a good capability for social learning. Meanwhile, there are some other sociable robots, which were designed to establish relationships with humans, for example, Autom (Kidd and Breazeal, 2008), which was designed to play a role as a weight loss coach to guide humans to loss their weight, which showed that developing long-term close relationship between human and robot was practicable, ATR's Robovie (Kahn et al. 2008), which had a social characteristic that could provide physical intimacy (e.g., holding, touching, embracing) with human, and MAGGIE (Gorostiza et al. 2006), whose main goal is to interact in a natural way and establish a peer-to-peer relationship with humans. Moreover, there was a research platform for studying social robotic user-interfaces, called iCat ("interactive cat"), developed by Philips, which was equipped with the capabilities of vision, conversation and facial expression and used to examine the influence of the perceived social abilities of a robot on user's attitude towards and acceptance of the robot (Heerink et al. 2006). And in (Heerink et al. 2006), researchers found out that elders are more comfortable with a more sociable robot, which indicated that elderly users were generally comfortable in communicating with the iCat interface.

The above sociable robots were equipped with ability to express emotions, the capabilities of which can be the robot attributes for social interaction. However, there are some other robot attributes. The main topics explored in robot attributes include robot's appearance, such as Uncanny Valley (Mori, 1970; Walters et al. 2008; Kanda et al. 2008; Ishiguro, 2006) and Social Space (Walters et al. 2006; Walters, 2008), robot's behavior such as Turn-Taking (Thórisson, 2002; Calinon and Billard, 2006; Kose-Bagci et al. 2008; Breazeal, 2003b) and Contingency (Dautenhahn and Werry, 2000; Yamaoka et al. 2007), the synchronization of robot's capabilities (Marek et al. 2007; Katsushi et al. 2008; Tetsuya et al. 2008; Michalowski et al. 2008; Aucouturier et al. 2008;

Kosuge et al. 2008; Breazeal, 2001; Wahlster, 2003; Alexandersson et al. 2004; Minoru et al. 2006; Tatsuya et al. 2008), and moreover, Affective Computing (Picard, 1995; 1999; 2003; Tao et al. 2005; Pantic and Rothkrantz, 2003; Gunes et al. 2008; Cohn, 2010; Cohn & Sayette, 2010).

3.3.2 User Attributes

Regarding the user attributes, the results of recent experiments have shown that user attributes affect human-robot social interaction. In 2008, 16 Chinese and 16 German college students participated in Rau et al. (2009)'s laboratory experiment in China, the object of which was to investigate the effects of communication styles and culture on people's tendency to accept recommendations from robots. The result of the experiment showed that there was a cultural difference in the participants' evaluations of the robot and their acceptance of the robot's recommendations, and that the subjects preferred the robot to communicate in the interpersonal communication style familiar to them. As compared with German participants, the Chinese participants preferred an implicit communication style, and were more likely to accept the implicit recommendations. Robots with an implicit communication style were evaluated by them as being more likable, trustworthy, and credible, whereas the German participants evaluated the robots as being less likable, trustworthy, and credible, and were less inclined to accept implicit recommendations. Moreover, by examining the effects of a sad mood on human facial emotion recognition in Chinese people, researchers (Lee et al. 2008) in Hong Kong found out that when the targets were ambiguous, participants who were in a sad mood tended to classify them in the negative emotional categories rather than the positive emotional categories, which indicated that emotion-specific negative bias in the judgment of human facial expressions is associated with a sad mood. Burleson and Picard (2007) conducted an experiment which 11-13-year olds were helped by an emotionally intelligent embodied agent to solve the Towers of Hanoi problem, the result of which turned out to be that: during the task, girls preferred an agent that provided affective support over one that provided support for the task itself, while boys preferred the agent that provided support for the task over the agent that displayed affective support, which indicated that the gender of a user could potentially have strong influence on perception of emotional expressions. The results of the above experiments suggest that when a researcher designs a human-robot interaction, special attention should be paid to effects of communication style, sad mood, and gender of a user on perception of emotional expressions of the sociable robots.

3.3.3 Task Structure

Since it can be a key element of human-robot interaction design, the effects of task structure on users' perception are also worth probing. Mutlu et al. (2006) used the humanoid robot-Honda's ASIMO to experimentally demonstrate effects of user attribute of gender and the task structure of competition and cooperation on perception of the emotional expressions of the sociable robots. They found out that men perceived the robot ASIMO less positively in the competitive task than in the cooperative task, while women's perceptions relied on the social attributes of the robot and did not change based on the task structure, which suggested that the task structure and the individual attributes of users should be matched to the interaction style of the robot during the designing of human-robot interaction.

3.4 Summary and Conclusion

This chapter contains a review of how different theories of emotion in psychology are adapted to create convincing and believable synthetic emotions of some emotional robots. In addition, androids such as Bristol Robotics Laboratory's Jules (Jaeckel et al. 2008a; 2008b; Henrik et al. 2009) using certain mathematical models which map from facial expressions of human beings to a robot's muscle or servo space are introduced.

Socially interactive robots, varying from humanoid, android, creature-like, to simply non-humanoid and non-zoomorphic in form, are often equipped with rich facial features to enable them to participate in emotion based interactions. The facial expressions of a robot can be created by means of the FACS (Ekman and Friesen, 1978; 2002). These robots should be able to only express a single affective state at a time as they can only be in a single state at a time (Breazeal, 2003a). They also need to be able to communicate this single affective state to humans through a discrete facial expression. Hence, the accuracy of selecting and conveying a robot's specific affective state, along with the timing of its emotional response, become important factors in achieving efficient communication. The arbitration of its specific affective state and the control of the timing can be processed by its own affect space, for instance, Russell's circumplex model of affect (Russell, 1997; Posner et al. 2005). As has been mentioned before, a robot's computational model of emotion becomes useful when coordinating a robot to respond to the external environment appropriately. The importance of such a model has been demonstrated in the Kismet project (Breazeal, 2002; 2003), which took a more complicated approach to a robot's emotions than that of the Probo project (Saldien et al. 2010).

However, the external environment that these robots interacted with in the previous experiments did not involve surrounding contexts with high emotional valence (e.g., valenced music or video). For instance, in the Kismet project, the effect of the external environment on the robot's internal state, which in turn affects its emotional expressions, was focused on. It seems that researchers in this field have made little effort to match the robot's emotional responses to the emotional external environment (e.g., a surrounding emotional context).

Moreover, Creed and Beale (2005) pointed out that little is known about how humans respond psychologically to synthetic emotions (through the use of textual content, speech (synthetic and recorded), and synthetic facial expressions) and what effect they have on a user's perceptions, behavior and performance. In their paper, they raised several questions for future experiments, such as “Can emotional agents influence or change a user's attitudes, beliefs, and behavior more effectively than unemotional agents?”. Furthermore, Beale and Creed (2009) stated that “*it is not sufficient to simply ask whether emotional agents are better or worse than unemotional agents. Instead, a better question is that of which kind of emotional expression, expressed in which way, influences which elements of a user's perceptions and behavior?*”. (Beale and Creed, 2009, p. 757)

There are some studies looking at Beale and Creed’s question. For example, Moshkina and Arkin’s (2005) experiment which involved a Sony robotic dog Aibo arrived at the conclusion that:(1) participants believed that the robot displayed emotions and/or personality and their interaction was made more pleasant by emotions and/or personality displayed by the robot at the same time; (2) the participants’ negative mood was reduced when they interacted with a robot that displayed emotions; (3) women were more susceptible to emotional cues and had more desire to show their emotions to the robot. The above results also supported the hypothesis that emotional agents influence or change a user's attitudes, beliefs, and behavior more effectively than unemotional agents. Regarding the iCat, researchers have not only used it to examine the influence of perceived social abilities of a robot on a user’s attitude towards and acceptance of the robot (Heerink et al. 2006), but also used it to examine the effects of iCat’s emotional behavior on users’ perception (Leite et al. 2008). Leite et al. (2008) used the iCat robot and chess (with physical electronic chessboard that detects the board state and sends it to the computer) to develop a chess game platform, where the robot’s affective state was influenced by every move the player made, which meant that users could know whether they had made a good or a bad move by looking at iCat’s expressions. By investigating the effects of the robot’s emotional behavior on the user’s perception of the game state, they found that a social robot with emotional behavior could perform the task of helping users to understand a gaming situation better than a robot without emotional behavior.

Empirical research involving the Sony robotic dog Aibo and Philips emotional robot iCat are not the only studies investigating the significance of user attributes (e.g., attitudes, beliefs, and behavior) besides robot attributes (e.g., robot’s emotional capability). Beale and Creed (2009) pointed out that a number of recent studies have examined the influence of synthetic agent emotion on user attitudes, perceptions and behavior (for examples, Maldonado et al. 2005; Prendinger et al. 2003; Berry et al. 2005; Morkes et al. 1998). Although the results of these studies were often inconclusive and contradictory, it was clear from all of these studies that synthetic emotions expressed by agents had the potential to influence user attitudes and behavior in a variety of ways, even though they were only rated as having a negative impact on the interaction on a couple of occasions.

As Beale and Creed (2009) suggested, it is worth exploring the question of which kind of emotional expression, expressed in which way, influences which elements of a user’s perceptions and behavior. As can be seen in the cases of Kismet (Breazeal, 2002) and Probo (Saldien et al. 2010), the facial expressions of the robot were only perceived in the absence of any surrounding emotional context (e.g., music or video). Will participants’ judgments of the facial cues be different given a surrounding emotional context? Will the mood states of the participants be affected by the facial cues of the robot or by the surrounding emotional context?

Beale and Creed (2009) compiled the following suggestions for investigations of the ways in which emotional agents affect users:

(1) Always validate emotional expressions: Before conducting an experiment, one should make sure that subjects can perceive the emotional expressions from an emotional agent relatively correctly. The extent to which users perceive the emotional expressions as the researchers intended should be tested as a pre-experiment.

(2) Fair comparison of emotion: One should ensure that the comparison between conditions is fair, for example, an emotional agent versus an unemotional one, a happy and warm emotional agent versus a sad and cold emotional agent, and strictly controlled experimental scenario versus less controlled and more natural environment. Failure to do this ultimately results in the validity of the reported effects being lowered.

(3) Be explicit about descriptions of emotional expression: In order for other researchers to assess the validity of experimental results, it is imperative that authors go into detail about which emotions were expressed, when they were expressed (e.g. what events triggered a response), which (if any) model of emotion (basic emotions versus cognitive emotions such as frustration, grief, humiliation, pride, and jealousy) was used, and how they were expressed (i.e. what was said, what textual content was used, what facial expressions were used).

(4) Statistics: All means and standard deviations should be included in the paper to facilitate statistical comparisons between studies.

(5) Finer grained approach: Researchers should focus on exploring which types of emotions, expressed in which ways, have which type of influence on an interaction.

The above suggestions serve as my guidelines for the experiments reported in this thesis. A set of experiments, partially designed to answer Beale and Creed (2009)'s questions about which kind of emotional expression, expressed in which way, influences which elements of a user's perceptions and behavior, are presented in next chapter.

Chapter 4: Contextual Effects on Judgments about Synthetic Robot Facial Expressions

As can be seen in Chapter 3, previous emotional robotic projects, especially some non-humanoid robotic projects, such as Feelix (Canamero and Fredslund, 2000), Kismet (Breazeal, 2002), Aryan (Mobahi et al. 2003), EDDIE (Sosnowski et al. 2006), and Probo (Saldien et al. 2010), have focused on using the FACS (Facial Action Coding System) (Ekman and Friesen, 1978 and 2002) and adapting Russell's circumplex model (Russell, 1997; Posner et al. 2005), to create synthetic robot facial expressions that are quite well recognised. However, so far none of these robotic projects have empirically investigated whether or not the surrounding context can also affect how humans perceive synthetic robot facial expressions. Therefore, this unanswered question will be explored in this chapter with 11 between-subjects experiments.

4.1 Generic Experimental Method

4.1.1 The Motivation of This Thesis

As reviewed in Chapter 2, some of the theories of emotion are useful in accounting for how humans recognize the emotional expressions of other people. On the basis of these theories it is possible to make predictions about the effect of a non-neutral contextual environment on the emotional judgments about robotic facial expressions. The predictions differ depending on the theory. According to Evolutionary Theories, there will be no contextual effect even in an emotional contextual environment - the message of a robotic EE (Expression of Emotion) encoded by the programmer is the one decoded by the receiver (human subject). This is because, most of the convincing and believable robotic facial expressions were created with the help of the FACS, which rested on the assumptions drawn from Evolutionary Theories that the same message of an EE (Expression of Emotion) encoded by the sender (or expresser) is the one decoded by the receiver (or observer).

In contrast, Social Constructionist Theories strongly predict that the situational information will influence the recognition of synthetic robot facial expressions: even when they are based on the FACS system the message of a robotic EE (Expression of Emotion) encoded by the programmer is not always the same as that decoded by the receiver (human subject). This prediction is actually supported by an argument in (Russell et al. 2003) that, attributing a specific emotion to the sender becomes more complex (i.e., the receiver does not simply decode an emotional message in any simple, reflex-like manner), as even when the stimuli are facial expressions of basic emotions, the context which they occur (Carroll and Russell, 1996) can color the attribution. Furthermore, some non-classical theories of emotion also predict, though not strongly, there will be some sort of contextual effect on the recognition of robotic facial expressions. For instance, according to Niedenthal's theory of embodying emotion (Niedenthal et al. 2007), emotion congruence in perception strongly affects later reactivation and recognition of a facial expression of an emotion in another person. This argument predicts that congruent emotional contexts will help human subjects to recognize the robotic facial expressions better than the incongruent contexts. Even Perceptual Control Theory (PCT) claims that the disturbance from the environment could directly elicit an emotion of a living organism, which can lead to the prediction that a context can affect a person's emotional states and his/her appraisals of a robot's emotions.

As reviewed in Chapter 2, both the Embodied Appraisal Theory (e.g., Prinz, 2004a, 2004b) and the Social Functionalist Approach to Emotions (e.g., Keltner et al. 2006) claim that human emotions are both biologically based and socially constructed. Their concessions to Social Constructionist Theories enable them to predict that the contexts will influence the recognition of synthetic robot facial expressions to some extent. However, no position is taken with regards to situational influence in Cognitive-appraisal Theories or Rolls's Theory of Emotion, as it is too difficult for them to take into account the social context in which the emotions are often embedded.

Some implications about which predictions were correct will be yielded from the experiments reported in this thesis, depending on whether or not a non-neutral contextual environment is found to have an influence on emotional judgments about robotic facial expressions. If no contextual influence on robot expressions is found, this would imply that the prediction made by the Evolutionary Theories was right and the one made by Social Constructionist Theories was wrong. However, if a contextual effect on robot expressions is found, this implies that the prediction made by Social Constructionist Theories was right but the one made by Evolutionary Theories was wrong. Although the predictions made by Niedenthal's theory of embodying emotion and Perceptual Control Theory (PCT) were not as strong as those of the Social Constructionist Theories, they would also be implied to be correct if a contextual influence was found. Some implications about whether or not robot faces are viewed in the same way as human faces could also be drawn depending on whether or not a contextual effect was found on robot expressions. If no contextual effect was found, this would indicate that the robot emotions were not socially constructed, then the implications is that robot faces are not viewed in the same way as human faces. This is because, as mentioned earlier, human emotions are both biologically based and socially constructed. In contrast, the occurrence of contextual effects would imply that human beings view the robot face in the same way as they view human faces.

The main motivation of the thesis is to see whether the context effects that have been obtained with human faces and avatars also apply to robot faces, and to investigate whether all forms of context also exert such an effect. To see the circumstances under which such contextual effects are likely to occur, investigations will also be conducted to establish whether the manner in which the context and the robot head are presented (simultaneously or separately) is important in obtaining such a contextual effect. In other words, if context effects are found then the further questions to be investigated are (i) what kinds of context exert an effect? (ii) do contextual effects still occur when subjects have seen the robot face before? (iii) does the manner in which the context and the robot face are presented (simultaneously or separately) make a difference? Where possible, the experimenter is also interested in exploring and differentiating between possible explanations of the reason context effects occur and do not occur, but it should be noted that this is not the preliminary motivation of the experiments reported here.

4.1.2 The Hypotheses

Earlier studies with human faces (e.g., Carroll and Russell, 1996; De Gelder and Vroomen, 2000) and avatar faces (e.g., Mower et al. 2008; 2009) have shown that the way in which facial and contextual cues occurred (congruent or incongruent with each other), affected observers' judgments about the facial expressions of humans or avatars. Contextual effects were found in these studies: when confronted with a congruent combination of facial expression and a context, participants recognized the facial expression better than when they were presented with a conflicting combination of the expression and the context. In order to examine whether such effects will also be obtained in studies with robot faces, the methodology used in these human and avatar studies was used in selecting the way in which the facial cues of a robot and the contextual cues were presented. Hypothesis 1 (H1) stated that subjects will recognize the sequence of robotic facial expressions better in a congruent context than in a conflicting context.

Also, to look at the relative weight of the surrounding context and the robot face on the emotional judgments of the robot face, hypothesis 2 (H2) was formulated that when the emotional valences of the context and the robot face are conflicting, the context will have a more dominant effect over the robot face in attributing the robot's facial expressions to its emotions. This hypothesis was based on the findings that a stronger effect of surrounding context was found on the recognition of avatar and human faces (Carroll and Russell, 1996; Noël et al. 2009; Mower et al. 2008; Mower et al. 2009) while a bidirectional effect of a surrounding context was found in earlier studies (De Gelder and Vroomen, 2000; Hong et al. 2002; Creed and Beale, 2008) with avatars or human beings.

4.1.3 The Dependent Variable

In each experiment, subjects' recognition of the positive/negative sequence of simulated facial expressions of a robot was the dependent variable. Two sequences of simulated facial expressions were developed to be shown either with accompanying emotional contexts or after the preceding emotional contexts. The sequences consisted of two types: Positive Affect, which mainly consisted of three different versions of joyful and surprised expressions (motions such as looking around and nodding were added in the gaps between these joyful and surprised expressions), and Negative Affect which mainly consisted of three different versions of sad, angry, and disgusted expressions (motions such as shaking and denying were added in the gaps between these sad, angry, and disgusted expressions). Each sequence of facial expressions was about three minutes long: almost the same time length as each type of the surrounding contexts.

These readable facial expressions of a robot face were created based on the FACS (Facial Action Coding System) (Ekman and Friesen, 2002). Ekman's Big Six emotions (happiness, sadness, fear, surprise, anger and disgust) were chosen to be expressed by the robot head, based on the fact that the Big Six emotions were widely chosen in most of the previous projects involving robotic facial expressions. The 11 experiments reported here were based on a robot head known as CIM, a believable emotional agent without vision and audio sensing abilities in the NRG lab, the University of Sheffield (see (Bailey, 2006) for more details). The electronic interface with the CIM robot has been built around the Sun SPOT system incorporating an I2C bus to connect the individual components to the dedicated microprocessor. This microprocessor is a dedicated java processor capable of running java programs that conform to a cut down definition of the full Java 1.1.8 API. The green cameras on either side of the head are designed to be blob cameras.

CIM is equipped with the following functional components and the movement options available to each component:

- Left eye and right eye– Yaw and Tilt abilities
- Left eyebrow and right eyebrow– Capable of raising on the right and left sides
- Upper and lower lip– Capable of bending up and down
- Neck– Allowing the head to turn and tilt
- Nose– Allowing movement backward and forward
- Cheeks– Capable of rotating up and down
- Iris control– Allowing modulation of how much light shines through the Pupils

Despite the fact that CIM has an admittedly limited ability to articulate expressions - it lacks skin and hair and has limited degrees of freedom, the FACS was successfully applied to set up the parameters of the servos (see Table 4.1) to make the robot head produce sequences of the six static facial expressions (joy (e.g., AUs 6+12), fear (e.g., AUs 1+2+4+10+12), surprise (e.g., AUs 1+2+10+12+58+63), anger (e.g., AUs 2+23+42+44), disgust (e.g., AUs 4+9+15+16), and sadness (e.g., AUs 1+4+6+15+64)).

Table 4.1 Servos set up for CIM.

Facial expressions	Action set up	Servo Speed
Joy	Smiling Lips, Raised Checks	Relatively Fast
Sad	Crying Brows, Crying Lips, Raised Checks, Eyes Down	Slowest
Anger	Angry Brows, Eyes Narrowed, Tightened Lips	Fast
Fear	Raised Eye Brows, Mouth Opened	Fast
Disgust	Frown Eye Brows, Nose Wrinkled, Raised Upper Lip	Slow
Surprise	Raised Eye Brows, Eyes Widened, Eyes Up, Neck Backward	Fastest

Examples of simulated facial expressions shown by CIM (see Fig. 4.1) are presented as follows:

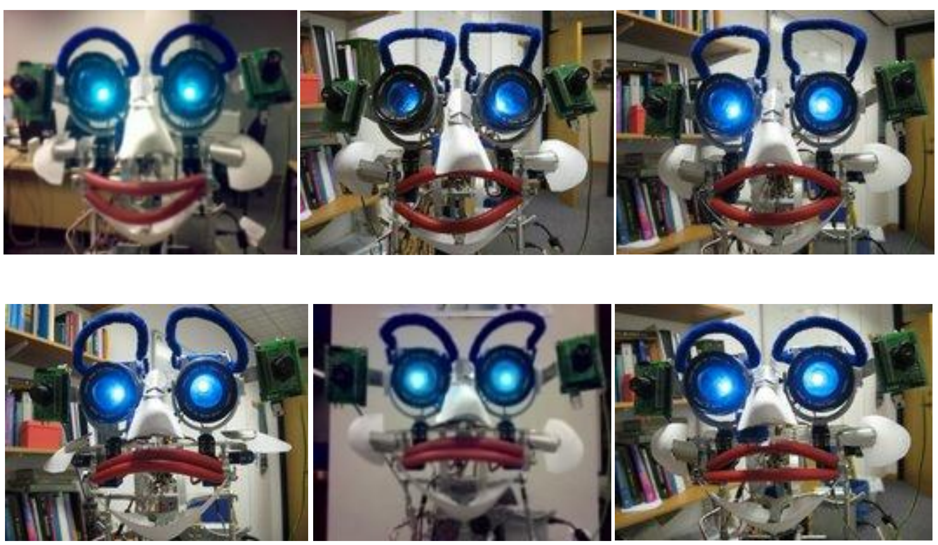


Fig. 4.1: Joy (top left), Surprise (top middle), Fear (top right), Sadness (bottom left), Anger (bottom middle), and Disgust (bottom right) of CIM. Note that the robot head neither have lid nor have jaw resulting in no AU 7 or AU 26, for instance, in use.

4.1.4 The Independent Variable

With respect to the independent variable, four different types of materials were chosen to be the surrounding contexts or preceding contexts, which served as the independent variables. These materials were recorded examples of BBC News, the selected affective pictures from the IAPS (Bradley and Lang, 2007), the classical music pieces selected based on a list provided in (Eich et al, 2007), the film clips selected based on a list provided in (Rottenberg et al. 2007). The News recordings were selected from the BBC World News, and the latter three materials were validated as successful materials in previous studies. Each type of the surrounding or preceding emotional contexts can be either positive or negative. To assess the emotions elicited by the contexts, the Brief Mood Introspection Scale (BMIS) (Mayer and Gaschke, 1988), the Affect Grid (Russell et al. 1989) and the self-assessment manikin (SAM; Lang, 1980) were chosen to self-report the emotional experience of each participant, due to the fact that these rating instruments are easy to access and use, and had been previously empirically validated for this purpose.

4.1.5 The Experimental Conditions

The experiments were mainly designed to find out if the independent variable (positive/negative surrounding or preceding emotional context) had an effect on the dependent variable (subjects' recognition of the positive/negative sequence of the robotic facial expressions), as stated in H1. Therefore, in all the experiments, subjects were divided into 4 groups, namely, Group 1: Positive Context (surrounding or preceding) with Positive Robotic Sequence; Group 2: Positive Context (surrounding or preceding) with Negative Robotic Sequence; Group 3: Negative Context (surrounding or preceding) with Positive Robotic Sequence; and Group 4: Negative Context (surrounding or preceding) with Negative Robotic Sequence. Among these groups, Group1 and Group 4 belong to the Congruent Condition, while Group 2 and Group 3 belong to the Incongruent Condition. According to H1, subjects in the Congruent Condition will recognize the sequence of the robotic facial expressions better than subjects in the Incongruent Condition. If H1 was validated, further investigation would establish whether H2, that in the Incongruent Condition subjects will recognize the context better than the sequence of the robotic facial expressions, was also validated. The grouping of the subjects is illustrated in Fig. 4.2 and Fig. 4.3 as follows:

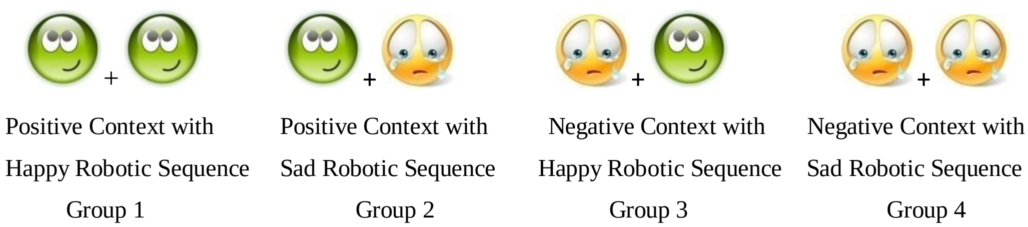


Fig. 4.2: Grouping of subjects with a surrounding context

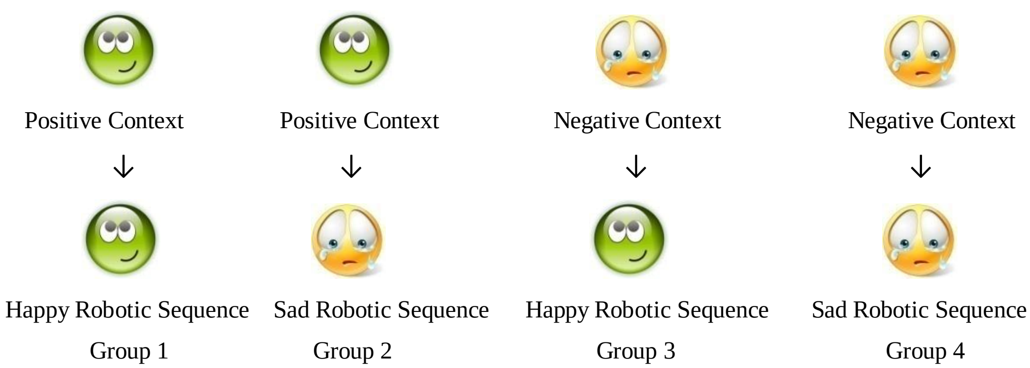


Fig. 4.3: Grouping of subjects with a preceding context

Most of the subjects recruited for the experiments were students in the University of Sheffield, and a small portion of them were university staff. Subjects were recruited from the university mailing list, or leaflets distributed to the locations such as the Student Union, some student accommodations in the university, and some lecture rooms in the university. In some experiments, subjects were not paid while in other experiments they were paid. Subjects belonged to various ethnic groups as the University of Sheffield has many international students. Subjects of both genders were recruited and there was no bias on the gender of the subjects. It is important to note that care was taken to ensure that subjects did not take part in more than one experimental session.

4.1.6 The Number of the Experiments

This chapter will describe 11 experiments (see Table 4.2) involving four different kinds of emotional materials, namely News recordings, affective pictures, musical pieces and film clips that were presented either simultaneously with a robot face or ahead of the presentation of a robot face.

Table 4.2 The 11 Experiments presented in this thesis

Context Types		Experiment	Manner Of Presentation (simultaneous/separate)	Viewing Order (first/second)
News Recordings		Speech1	Simultaneous	First
		Speech2	Simultaneous	Second
Affective Pictures		Image1	Simultaneous	First
		Image2	Simultaneous	Second
		Image3	Separate	First
Musical Pieces	(3mins)	Music1	Simultaneous	First
	(3mins)	Music2	Simultaneous	Second
	(20mins)	Music3	Separate	Second
Film Clips		Video1	Simultaneous	First
		Video2	Simultaneous	Second
		Video3	Separate	First

Note: 1. The same News recordings were used in Speech; the same affective pictures were used in Image; the same approximate 3mins long musical pieces were used in Music1 and Music2 while the longer version musical pieces were used in Music3; and the same film clips were used in Video. 2. In some cases, experiments were conducted in pairs as one experimental session (i.e., Image1 and Music3 in one session, Image3 and Video2 in one session, Music1 and Image2 in one session, Video1 and Music2 in one session, and Video3 and Speech2 in one session). This was done to save time and money when recruiting participants.

As can be seen from Table 4.2, the experimental results will be reported in an order according to the media type – speech audio, still image, music and video (a list of the actual order of the 11 experiments is provided in the appendix Ⅰ). This order allows the author to investigate one form of context at a time, and to establish which forms of context can exert a contextual effect. If a contextual effect was found for the speech audio, then further investigations would be carried out to see if three other types of emotional context, namely still image, music and video can also have contextual effects. The same sequences of synthetic robot facial expressions were presented to the subjects with one of the contexts at a time, presented either simultaneously with the context (see Fig. 4.2), or after the context (see Fig. 4.3). For each form of context, to see the circumstances under which such contextual effects are likely to occur, investigations were also undertaken to establish whether the fact that subjects have seen the robot face before or whether the manner in which the context and the robot head are presented (simultaneously or separately) makes a difference.

The idea that a second presentation of the robot face could weaken a contextual effect comes from a conclusion made by Niedenthal et al. (2006b) that previously encountered facial stimuli can color the recognition of facial expressions of human beings, as mentioned in Chapter 1. This is because such a contrast effect can not only increase the salience of the target robot facial expressions but also change the attribution of the emotion category label. In order to establish whether subjects seeing the robot face before makes a difference in obtaining a contextual effect, in some of the experiments (e.g., Speech2, Image2, Music2 and Video2), subjects saw the robot head and its sequence of expressions a second time (i.e., on a second viewing), accompanied or preceded by a different context. In such cases, a differently valenced sequence of robot expressions was shown: if they had seen a sequence of positive expressions, then the second viewing consisted of the robot head showing a sequence of negative expressions. As a result of the two oppositely valenced sequences of expressions, a contrast effect could have been generated. However, seeing whether a second presentation of the robot face would weaken any contextual effect found was not the main purpose of conducting experiments in pairs. The main purpose was to save time and money when recruiting participants.

As for the investigation of whether the manner in which the context and the robot face are presented (simultaneously or separately) makes a difference, this was included because of the likelihood that timing may play an important role in human's perception of synthetic robot facial expressions. Niedenthal et al. (2010) agreed with one commentator on their paper, Barlett (2010) who argued that timing principles are important in emotional information processing, as emotions are defined in part in terms of the timing of their onset and their duration (Tomkins, 1962; 1963). If the context under consideration is presented before the synthetic robot emotions are viewed, instead of simultaneously, the context may not exert the same effect as it does when the head and context are viewed simultaneously. For instance, after the experiments involving speech audio were conducted, further experiments were conducted to establish whether or not a contextual effect would also be found in a separate presentation of the context and the robot face (i.e., Image3, Music3 and Video3).

4.1.7 The Generic Experimental Procedures

In each experiment, subjects first watched/listened to the emotional material that was presented either simultaneously (as illustrated in Fig. 4.2) with or before (as illustrated in Fig. 4.3) the sequence of the robotic facial expressions. Then they answered a few questions to indicate how well they recognized the emotional content of the context or the robot expressions, such as follows:

1. As a total impression, please select what kind of material you think you were viewing/listening to from the following given choices. ___________

A: Positive Material   B: Neutral Material   C: Negative Material

2. As a total impression, please select what kind of emotion (affect) you think the robot was feeling from the following given choices. ___________

A: Positive Affect   B: Neutral Affect   C: Negative Affect

In some cases, subjects' mood states were measured only after the presentation of the emotional material without a pre-measurement before the presentation of the emotional material. For instance, in Speech1 (see Appendix Ⅲ: The Questionnaire of the Experiment Speech1), subjects' mood states were measured by the Brief Mood Introspection Scale (BMIS) after the context and robot head had been viewed. And in Image1, Image2, and Video1, subjects' mood states were measured by the Affect Grid only after the presentation of the emotional material. In other cases (i.e., Music2, Music3 and Video2), subjects filled in Affect Grids before the short questionnaire, once before and once after the presentation of the emotional context, to measure their mood states. However, only in Music2, Music3 and Video2 can the measurements of subjects' mood states be taken as a clear indication of the effects, or lack of effects, on subjects moods, as it was only in these conditions that the moods were measured twice. In experiments in which moods were only measured after the emotional material, it is possible that unmeasured pre-existing mood states could have affected any results obtained. Pre-existing mood states could have contributed to any effects on mood that were found, or could have masked (hidden) mood effects that would otherwise have been found. Therefore, subjects' mood states measured once without a pre-measurement will be reported and discussed in only one case – Experiment Speech1, while more time will be devoted to report and discuss the mood states of the subjects in the three experiments - Music2, Music3 and Video2 in which the moods were measured twice.

For more details about the experimental set ups, please refer to Appendix Ⅱ: Detailed Experimental Set Ups for All the 11 Experiments.

For detailed experimental procedures of each experiment, please refer to Appendix Ⅲ: The Questionnaire of the Experiment Speech1; Appendix Ⅳ: The Procedures of Experiments Image1 and Music3 Conducted in Pairs; Appendix Ⅴ: The Procedures of Experiments Image3 and Video2 Conducted in Pairs; Appendix Ⅵ: The Procedures of Experiments Video1 and Music2 Conducted in Pairs; Appendix Ⅶ: The Procedures of Experiments Music1 and Image2 Conducted in Pairs; Appendix Ⅷ: The Procedures of Experiments Video3 and Speech2 Conducted in Pairs.

4.1.8 The Statistical Analysis Techniques

In this thesis all null hypotheses were rejected at the 5% level of significance, and the Fisher Exact Test was chosen as the main statistical analysis technique to test H1 and H2. Where subjects' mood states were reported, a two-way ANOVA was chosen to analyze whether the emotional context or the robotic facial expressions had a mood effect on the subjects. The choice of statistic can be explained as follows:

First of all, according to Greene and d'Oliveira (1999), there are differences in the assumptions made by the T-test, the ANOVA test, and the Chi-Square, along with Fisher's Exact Test. For instance, T-test and ANOVA compare the difference between conditions, and are examples of parametric tests that have three prerequisites. First of all, that the data is interval data which is suitable for numerical calculations; secondly, the scores are normally distributed; and thirdly, variances in conditions are approximately the same. By contrast, Chi-Square, and Fisher Exact Test deal with the relationships between variables represented by nominal categories. And both of them are non- parametric tests. As both H1 and H1 examine the relationships between the dependent variable (subjects' recognition of the positive/negative sequence of simulated facial expressions of a robot) and the independent variable (four different types of emotional material) represented by nominal categories, the choice of statistic goes to Chi-Square or Fisher Exact Test. And since subjects' mood states were measured in Affect Grid, which were represented by interval numbers and satisfied the three prerequisites for parametric tests, a two-way ANOVA rather than a T-test was chosen to test the moods as there were two variables.

As for the choice of Chi-Square or Fisher Exact Test, it should be pointed out that Chi-Square with Yates' correction was used as the main statistical analysis technique in the two conference papers and the journal paper published by this author, but in this thesis it was used only as an auxiliary technique. The reason for this is, as stated on the website (http://www.graphpad.com/quickcalcs/contingency1.cfm), that although there are three ways to compute a P value from a contingency table, Fisher's test is the best choice as it always gives the exact *p* value, while the Chi-Square test only calculates an approximate *p* value. The Yate's continuity correction is to make the Chi-Square approximate better. With large sample size, the Yate's continuity correction makes little difference. With small sample size, the Chi-Square test is not accurate, with or without the Yate's continuity correction.

4.2 Experiment Speech1

Context: recorded BBC News

Manner of Presentation: simultaneously presented with the robot's expressions

Viewing Order of the Robot Face: first viewing

Mood States of the Subjects: not measured

Subjects were divided into four groups according to different combinations of BBC News (positive/neutral vs. negative) and robot expressions (positive vs. negative) in the manner as described in section 4.1.5.

The 72 subjects (38 male and 34 female) with average age 26.82 who participated in this experiment, had various nationalities (41 British, 4 Chinese, 8 other Asian participants, and 13 other European participants). They were recruited from the university mailing list, and they were not paid, but were offered the chance to win a book token worth 45-pounds. There were 18 subjects in each group, which meant that there were 36 subjects in the Congruent Condition (group 1 and group 4) and 36 subjects in the Incongruent Condition (group 2 and group 3).

The source of the recorded News that the subjects listened to was BBC World News (see Table 4.3). Two types of News were used in this study, one was Positive/Neutral News, and the other one was Negative News.

Table 4.3 Content of the recorded BBC World News.

No.\Type	Positive/Neutral News (date)	Negative News (date)
1	World's oldest spider web was discovered in Britain *(01/11/2009).*	Collective international action was called to prevent the tiger population from dying out *(28/10/2009).*
2	NASA said a 'significant amount' of frozen water has been found on the moon *(14/11/2009).*	Scientists found that snow and ice on Africa's highest mountain Kilimanjaro was melting rapidly and could vanish within 20 years *(03/11/2009).*
3	Scientists identified more than 17,000 species in the deepest parts of the oceans *(23/11/2009).*	British scientists warned that the medical use in humans of nanotechnology may damage the DNA or genetic building blocks of cells *(06/11/2009).*
4	Scientists used the Large Hadron Collider (LHC) at the CERN to produce the first particle collisions *(24/11/2009).*	The United Nations children's agency said nearly 200 million children under the age of five living in the developing world were stunted as a result of malnutrition *(12/11/2009).*
5	The draw for 2010 Football World Cup in South Africa took place in Cape Town *(05/12/2009).*	The United Nations Food Agency appealed for more money to feed one billion hungry people *(19/11/2009).*

4.2.1 Experimental Results

In the warming-up procedure (see Appendix Ⅰ), some of the robot's static expressions were more easily recognised than others (see Table 4.4). Most of the subjects (88%) could easily recognize the robot's sad expression, and a majority were able to identify the 'joy' expression (64%), and the surprise expression (65%). Anger was recognised correctly by 57% of the subjects, but most of the subjects (76%) confused the robot's fearful expression with a surprised expression. Moreover, the disgust and anger expressions of the robot were easily confused with each other, and some of the subjects (less than one third) found it difficult to tell the difference between the joy and surprise expressions.

Table 4.4 Percentage of six possible labels (items in the first row) chosen to match the displayed facial expressions (items in the first column) in E1.

% match	Joy	Sadness	Anger	Fear	Disgust	Surprise	% correct
Joy	64	0	0	3	5	28	64
Sadness	0	88	1	3	8	0	88
Anger	0	4	57	14	24	1	57
Fear	11	0	0	0	13	76	0
Disgust	0	4	45	11	40	0	40
Surprise	14	0	0	18	3	65	65

The responses to the questionnaire administered after viewing the robot were analysed. As shown in Table 4.5, subjects' judgements about the robot emotions do seem to be affected by the accompanying News. A response was considered to be correct when the robot head was said to have shown Positive Affect in the Positive Affect Condition, or Negative Affect in the Negative Affect Condition (the neutral choice was counted as wrong in both conditions). When the robot's positive expressions were accompanied by "positive/neutral" News, they were correctly recognised as such 94.4% of the time, as compared to only 44.4% of the time when accompanied by "negative" News. Conversely, the robot's negative expressions were correctly recognised as such 94.4% of the time when paired with "negative" News, as opposed to 33.3% of the time when paired with "positive/neutral" News.

Although subjects were often not good at identifying the robot's expressions, they were much better when the robot's expressions were accompanied by a congruent context. A Fisher's exact test in PASW 18 with two-tailed $p<0.0001$, indicated a significant association between information style (congruent or conflicting) and accuracy of attributing robot emotions to its facial expressions. In other words, the accuracy of subjects' perception of robot's emotions was significantly different depending on whether the robot behaved congruently with the BBC News (a higher accuracy, 34/36 (94.4% correct, 5.6% incorrect)), or whether the robot behaved conflictingly with the BBC News (a much lower accuracy, 14/36 (38.9% correct, 61.1% incorrect)). In addition, a Chi-square test for independence (with Yates Continuity Correction) in PASW 18 stated χ^2 (1, n=72) =22.562, two-tailed $p<0.0001$, correlation coefficient phi=0.589 (large effect). Consequently, H1 that subjects will recognize the sequence of robotic facial expressions better in a congruent context than in a conflicting context, was supported.

Table 4.5 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Speech1.

%match		Positive Affect	Neutral Affect	Negative Affect	% correct
Positive Affect Condition	group 1 (congruent News)	94.4	5.6	0	94.4
	group 3 (conflicting News)	44.4	27.8	27.8	44.4
Negative Affect Condition	group 2 (conflicting News)	44.4	22.3	33.3	33.3
	group 4 (congruent News)	0	5.6	94.4	94.4

It seems that subjects' identification of the emotion conveyed by the News was good, and did not seem to be affected by whether or not the robot face displayed matching emotions (see Table 4.6). A response was considered to be correct when the News was said to have been Positive or Neutral in the Positive/Neutral BBC News Condition, or Negative in the Negative BBC News Condition (neutral News was not counted in this condition).

Table 4.6 Percentage of three possible labels (items in the first row) chosen to match the played sequence of BBC News items (positive/neutral or negative) in Speech1.

%match		Positive News	Neutral News	Negative News	% correct
Positive/Neutral BBC News Condition	group 1 (congruent robot)	77.8	22.2	0	100
	group 2 (conflicting robot)	33.3	66.7	0	100
Negative BBC News Condition	group 3 (conflicting robot)	5.6	11.1	83.3	83.3
	group 4 (congruent robot)	0	5.6	94.4	94.4

The hypothesis (H2) that when the emotional valences of the context and the robot face are conflicting, the context will have a more dominant effect over the robot face in attributing the robot's facial expressions to its emotions, was also tested. A Fisher's exact test in PASW 18 with two-tailed p <0.0001, indicated the BBC News modality had a stronger perceptual effect than the Robot Expression modality when subjects were presented with conflicting information (33/36 accuracy for BBC News VS. 14/36 accuracy for Robot Emotions). In addition, a Chi-square test for independence (with Yates Continuity Correction) in PASW 18 stated that χ^2 (1, n=72) =19.854, two-tailed p<0.0001, correlation coefficient phi=-0.554 (large effect). Consequently, H2 was supported.

4.2.2 Discussion and Conclusion

This experiment was designed to determine whether a speech-based surrounding emotional context (congruent, or incongruent) influenced users' perception of a robot's simulated emotional expressions. Hypothesis H1 that subjects will recognize the sequence of robotic facial expressions better in a congruent context than in a conflicting context, was supported here. This suggests that the recognition of robot emotions is strongly affected by the particular speech context investigated here, and that such emotions are more likely to be recognized when they are congruent with that context. The second hypothesis H2, that when the emotional valences of the context and the robot face are conflicting, the context will have a more dominant effect over the robot face in attributing the robot's facial expressions to its emotions, was also clearly validated in this experiment. In summary, when confronted with a congruent surrounding context (the recorded BBC News), people were more able to recognize the robot emotions as intended than when faced by an incongruent surrounding context. In addition it was found that the recorded BBC News had a more dominant effect on judgments about the robot's expressions than the content of the expressions themselves.

How similar are the facial expressions, and the context used here to those employed before? The expressions used in the human face studies were often of basic emotions, while computer avatar studies tended to use more cognitive ones, such as the warmth and concerned emotions used in Creed and Beale (2008)'s study. And emotional voices are often used as a context for both the human face and avatar face studies. The facial expressions of a robot investigated here were of basic emotions, which were created based on the FACS, and the context presented with them was verbal descriptions in human neutral tones. It seems the context investigated here is similar to those employed before in human face studies, and the facial expressions of the robot share a lot in common with those used in avatar studies as both of them are based on the FACS.

And how similar is the contextual effect found here with a robot head to the contextual effects found with human faces and avatar faces? As mentioned in Chapter 1, some contexts were found to influence not only the recognition accuracies of human facial expressions but also the reaction time for judgments about them. For instance, despite the facial dominance account of human emotions established by Izard (1971) and Ekman (1972), later studies by Watson (1972), by Wallbott (1988), and by Carrera-Levillain and Fernández-Dols (1994), showed that contextual effects on the recognition of human emotional expressions are stronger when the expressions are ambiguous or natural and less intense. And De Gelder and Vroomen (2000) presented a series of pictures of faces ranging from happy to sad, accompanied by a sad voice, a happy voice or no voice, to demonstrate that when the emotions matched, people's reaction time in judgments was faster than when they were mismatched. Context has also been found to affect human interpretation of the emotional expressions of computer avatars (Creed and Beale, 2008; Noël et al. 2009; Mower et al. 2008; 2009), whose faces were often paired with voices (artificial or human), as also mentioned in Chapter 1. The particular speech context investigated here exerted a contextual effect on the robot expressions, which implies that the contextual effect found here with a robot head is also similar to the contextual effects found in the human face or avatar face studies.

Human emotions are both biologically based and socially constructed, as claimed by both the Embodied Appraisal Theory (e.g., Prinz, 2004a, 2004b) and the Social Functionalist Approach to Emotions (e.g., Keltner et al. 2006). Will the results of this experiment suggest that robot faces are viewed in the same way as human faces? The FACS, which was initially developed as a set of guidelines for recognizing the facial expressions of humans, has previously been shown to be a reliable tool to create convincing and believable facial expressions for computer avatars and social robots. It seems that the principles of the FACS can be used to recognize both human faces and robot faces. A contextual effect was found for the speech audio on the robot expressions, which were also based on the FACS. Therefore, an implication of the present results is that the prediction made by the Social Constructionist Theories was correct and the one made by the Evolutionary Theories was wrong. It seems that the message of a robotic EE (Expression of Emotion) encoded by the programmer is not always the same as that decoded by the receiver (human subject), which indicates that robot emotions are also social constructed. And predictions made by Niedenthal's theory of embodying emotion and Perceptual Control Theory (PCT) were also implied to be correct as a contextual influence was found. Consequently, the robot face seems to have been viewed in the same way as human faces.

Robot faces may be viewed like human faces, possibly partially because recognizing the facial expressions of an emotional robot also involves embodiment (emotional states reactivation or reenactment). As described in Chapter 2, the theory of embodying emotion may provide an explanation of emotional congruence in perception of human emotions, as emotion congruence in perception strongly affects later reactivation and emotion recognition. From an emotional congruence perspective, an effect of emotional congruence in perception of robot facial expressions was found when subjects were confronted with either congruent or conflicting emotional cues from the context and the robot face in this experiment. As a result, the emotional congruence in this study may be similar to emotional congruence in perception of human emotions, meaning that robot faces are viewed like human faces. And the theory of embodying emotion may also provide an explanation of emotional congruence in perception of robot facial expressions, although no empirical research has focused on the validation of embodied recognition of robot facial expressions.

How should the effects of an accompanying emotional context on judgments about the robot's emotional expressions be explained? It is possible to identify four possible explanations. One is that from an emotional congruence perspective, recognizing the facial expressions of an emotional robot also involves embodiment, as argued above. The second one is that the context affects the emotional state of the observer, which in turn affects their perception of the robot. Niedenthal et al. (2006b) concluded that observers of human facial expressions were influenced by their own emotional state in their attributions of emotional states to the people they observed. The third one is that the observer interprets the robot's expressions as though the robot was capable of "listening" to the speech and responding to it accordingly. In this explanation, the observers' judgments are not determined by their own emotional state. The fourth explanation from a pattern recognition perspective is that the contextual effect might be just a cue integration issue - subjects may have evaluated both sources together and chosen the context to reinterpret the robotic expressions. This would create a result in which the recorded BBC News had a more dominant effect on judgments about the robot's expressions over the robot face itself such that subjects' reinterpretation of the robot expressions was susceptible to the interpretation of the News recording. And the data collected in this experiment seems to support this explanation as H2 was supported.

It was not expected that the recorded BBC News would color subjects' mood states according to the content of the News itself (see Table 4.3). However, a Two-Way ANOVA (unrelated) in SPSS 16 was conducted to explore the impact of the recorded BBC News and the synthetic robot emotions on subjects' moods, as measured by the Brief Mood Introspection Scale (BMIS) (Mayer and Gaschke, 1988). The interaction effect between recorded BBC News and the synthetic robot emotions on the subjects' moods was not statistically significant, $F(1,68)= 0.032$, $p=0.858$; in addition, neither the main effect for recorded BBC News ($F(1,68)= 0.013$, $p=0.910$) nor the main effect for synthetic robot emotions ($F(1,68)= 0.021$, $p=0.884$) reached statistical significance. There was no significant difference in scores for participants who listened to the positive/neutral News ($M=47.5278$, $SD=6.45196$), and participants who listened to the negative News ($M=47.7222$, $SD=7.76664$). And there was no significant difference in scores for participants who watched positive sequence of robot facial expressions ($M=47.7500$, $SD=5.47918$), and participants who watched negative sequence of robot facial expressions ($M=47.5000$, SD=8.48023). It seems that neither the recorded BBC News nor the synthetic robot emotions appeared to affect subjects' moods. However, since the BMIS was only presented at the end of the experiment, there is still the possibility that subjects' moods were affected, but that these effects were masked by their pre-existing moods. Since no data was collected that could clearly indicate a significant association between subjects' moods and their perception of the synthetic robot emotions, it is not possible to draw strong conclusions about whether or not context effects were mediated by subjects' moods. It is suggested therefore that the effect of the surrounding context observed in this experiment was only partially due to the observers' own emotional state coloring their perception of the robot.

Observers' witnessing the context and the robot at the same time, and interpreting the robot's expressions as if the robot was responding to the surrounding context, could also be one of the factors causing a contextual effect. When the context was negative, the robot's expressions were seen to be negative, even when on objective criteria (the underlying FACS coding) they should have been seen as positive. The reverse was also true (the robot's expressions were seen as positive when the surrounding context was positive, even when there were objective reasons to expect the expression to be seen as a negative one). The interesting implication of this is that the robot's expressions seem to be being interpreted as if it were capable of understanding and responding to the current emotional situation.

Which explanation is better supported by the experiment reported here? The data collected in this experiment only relates to two of the explanations – the mood effect explanation and the cue integration explanation. And it would be too difficult to conduct experiments to validate the speculations about the remaining two explanations – the emotional embodiment explanation and the robot's responding to the context explanation. The present experiment does not make it possible to distinguish between the four explanations, or to eliminate any of them. It is possible that in this experiment, judgments about the robot's expressions were influenced by all four factors at the same time. More time will be devoted to discuss the mood effect explanation and the cue integration explanation, perhaps after the substantial data of a series of experiments is collected in the last session of the chapter.

Although it is difficult to distinguish between possible explanations of the contextual effect found here, the results of this experiment do indicate that the attribution of robotic facial expressions to its emotions can be affected by the situation in which they occur. In particular, a surrounding context of recorded News items containing content that was either positive/neutral, or negative, was found to affect the attribution of robotic facial expressions to its emotions. Distinguishing between possible explanations of a contextual effect is the not the main purpose of this thesis, but this discussion provides a first stab at analyzing the potential factors contributing to the contextual effect that found here. Further discussion of the possible explanations will be returned to the last session of this chapter.

This experiment establishes that a particular form of context, spoken News recordings, has the ability to exert an effect on the interpretation of a robot's emotional expressions. As mentioned earlier, when a contextual effect was found, further investigation will be carried out to discover the circumstances under which an emotional context will affect people's view of the robot's expressions. More specifically, the next experiment investigates whether subjects have seen the robot face before makes a difference in obtaining a contextual effect. Therefore, the next experiment, Speech2 was conducted to see if a second viewing of the robot face would weaken, or eliminate the any contextual effect.

4.3 Experiment Speech2

Context: recorded BBC News

Manner of Presentation: simultaneously presented with the robot's expressions

Viewing Order of the Robot Face: second viewing

Mood States of the Subjects: not measured

Subjects were again divided into four groups according to different combinations of BBC News (positive/neutral vs. negative) and robot expressions (positive vs. negative) in the manner as described in section 4.1.5.

52 volunteers with an average age of 23.23 participated in this experiment. These 31 male and 21 female participants had various nationalities (19 British, 12 Indian, 8 Chinese, 7 other Asian participants, and 3 other European participants). They were recruited from leaflets distributed to the locations such as the Student Union, some student accommodations in the university, and some lecture rooms in the university. Each participant was paid £5 for each taking part, as was advertised on every leaflet. There were 13 subjects in each group, resulting in 26 subjects in the Congruent Condition (group 1 and group 4), and 26 subjects in the Conflicting Condition (group 2 and group 3). Care was taken to ensure that no subjects had participated in the previous experimental sessions.

4.3.1 Experimental Results

The responses to the questionnaire administered after viewing the robot were analyzed. As shown in Table 4.7, subjects' judgments about the robot emotions do again seem to be affected by the accompanying News. A response was considered to be correct when the robot head was said to have shown Positive Affect in the Positive Affect Condition, or Negative Affect in the Negative Affect Condition (the neutral choice was counted as wrong in both conditions). When the robot's positive expressions were accompanied by "positive/neutral" News, they were correctly recognized as such 100% of the time, as compared to only 30.8% of the time when accompanied by "negative" News. Conversely, the robot's negative expressions were correctly recognized as such 76.9% of the time when paired with "negative" News, as opposed to 61.5% of the time when paired with "positive/neutral" News.

Moreover, statistical analysis showed these differences were significant. A Fisher's exact test in PASW 18 with two-tailed p=0.003 indicated a significant association between Information Style (Conflicting Information or Congruent Information) and Accuracy of recognizing robot's emotions, in other words, accuracy of subjects' perception of robot's emotions was significantly different between when the robot behaved congruent with recorded BBC News (a very high accuracy, 23/26 (88.5% correct, 11.5% incorrect)) , and when the robot behaved conflictingly with recorded BBC News (a very low accuracy, 12/26 (46.2% correct, 53.8% incorrect)). In addition, a Chi-square test for independence (with Yates Continuity Correction) in PASW 18 stated that $\chi 2$ (1, n=52) =8.739, two-tailed p=0.003 (to be significant, this should be smaller than 0.05), correlation coefficient phi=-0.451 (medium effect). Consequently, H1 that subjects will recognize the sequence of robotic facial expressions better in a congruent context than in a conflicting context, was supported. Note that in this experiment, subjects watched the synthetic robot emotions while listening to the recorded BBC News simultaneously, having seen the opposite valence of the robot emotions before the presentation of the recorded BBC News.

Table 4.7 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Speech2.

%match		Positive Affect	Neutral Affect	Negative Affect	% correct
Positive Affect Condition	group 1 (congruent News)	100	0	0	100
	group 3 (conflicting News)	30.8	23.1	46.1	30.8
Negative Affect Condition	group 2 (conflicting News)	7.7	30.8	61.5	61.5
	group 4 (congruent News)	7.7	15.4	76.9	76.9

It seems that subjects' identification of the emotion conveyed by the News was good, and did not seem to be affected by whether or not the robot face displayed matching emotions (see Table 4.8). A response was considered to be correct when the News was said to have been Positive or Neutral in the Positive/Neutral BBC News Condition, or Negative in the Negative BBC News Condition (neutral News was not counted in this condition).

Table 4.8 Percentage of three possible labels (items in the first row) chosen to match the played sequence of BBC News items (positive/neutral or negative) in Speech2.

%match		Positive News	Neutral News	Negative News	% correct
Positive/Neutral BBC News Condition	group 1 (congruent robot)	61.5	38.5	0	100
	group 2 (conflicting robot)	30.8	46.1	23.1	76.9
Negative BBC News Condition	group 3 (conflicting robot)	0	0	100	100
	group 4 (congruent robot)	0	15.4	84.6	84.6

The hypothesis (H2) that when the emotional valences of the context and the robot face are conflicting, the context will have a more dominant effect over the robot face in attributing the robot's facial expressions to its emotions, was also tested. A Fisher's exact test in PASW 18 with two-tailed p=0.003 indicated the recorded BBC News had a more dominant effect on the judgments about the emotions of the robot than the robot emotions themselves (23/26 accuracy for BBC News VS. 12/26 accuracy for Robot Emotions). In addition, a Chi-square test for independence (with Yates Continuity Correction) in PASW 18 stated that $\chi 2$ (1, n=52) =8.739, two-tailed p=0.003, correlation coefficient phi=-0.451 (medium effect). Consequently, H2 was supported in this case. Note that the p value in Speech1 (robot emotions were on first viewing) is less than 0.0001 which is smaller than the value obtained in this case. It seems that in the case of second viewing, the contrast effect turned out to weaken the contextual effect here such that the surrounding BBC News became less dominant on the emotional judgments of robot emotions.

4.3.2 Discussion and Conclusion

H1 and H2 were validated, showing that a contextual effect was still found despite the fact that the robot face and its expressions were being viewed for the second time. It indicated that a second viewing of the robot face did not weaken the contextual effects. The procedure and materials used in this experiment was identical to those used in Speech1, but subjects in Speech2 were new and this was their second viewing of a sequence of robot expressions. Even so, the particular speech audio investigated here was still found to have a dominant effect over the robot face in attributing its facial expressions to its emotions.

More experimentation is needed to explore the relative effects of different kinds of surrounding context. Nonetheless, it remains the case that the results indicate that the recognition of a robot's emotional expressions can be affected by a surrounding speech audio. Effectively the same pattern of results was obtained in experiment Speech1 as in experiment Speech2, even though the robot's face and its expressions were being viewed for the second time in experiment Speech2. The second viewing might have weakened the context effects – the fact that it does not seem to have done so suggests that the effect is reasonably robust.

Will the conclusions in Speech1 and Speech2 apply to other emotional materials such as still images, music and video? To find out if a contextual effect will be found on other emotional materials, it was necessary to replace the speech audio by a new media type, in this case, still images. Apart from changing the context, from an auditory one to a visual one, the same methodology will be applied in the next experiment.

The changes in the experimental set up for the Image experiments have some further implications for possible explanations of any context effects that may be found. These will be discussed later, but are mentioned here in order to highlight them. First, as mentioned in Chapter 2, for positive and negative concepts, verifying properties from different modalities produced costs such that reaction times were longer and error rates were higher than if no modality switching was required (Vermeulen et al 2007). Therefore, perceiving emotional cues from the same channels (e.g., visual channel (still images) with visual channel (robot face) in next experiment) should be easier than perceiving them from two channels (e.g., visual-audio in Speech1 and Speech2). As a contextual effect was already found for the speech audio, it could be expected that a contextual effect will be found for still images in next experiment. In addition, in the following experiments using images, the still images will be presented in such a way that the subjects should not think that the robot was capable of "seeing" the images. The images were presented on a computer screen which is parallel to the robot head. If a contextual effect is found for such still images, further studies will be undertaken to determine the circumstances under which such an emotional contextual affect people's view of the robot's expressions.

4.4 Experiment Image1

Context: affective pictures

Manner of Presentation: simultaneously presented with the robot's expressions

Viewing Order of the Robot Face: first viewing

Mood States of the Subjects: not measured

Subjects were again divided into four groups according to different combinations of affective pictures (positive vs. negative) and robot expressions (positive vs. negative) in the manner as described in section 4.1.5.

The affective pictures used in this experiment were selected from the international affective picture system (IAPS), as the IAPS is currently used in experimental investigations of emotion and attention worldwide, providing experimental control in selection of emotional stimuli (Bradley and Lang, 2007). The selected affective pictures were either all pleasant (No. 1463 1610 1710 1920 2040 2057 2070 2080 2150 2160 2311 2340 2360 2530 2550 2660 4220 4250 5480 5760 5910 7330 7350 7400 7470 7580 8190 8200 8210 8370 8470 8540) or all unpleasant (No. 2205 2590 2661 2710 2750 2800 2900 3180 3220 3230 3300 3350 3530 6260 6350 6560 6570 9040 9050 9110 9120 9181 9220 9330 9340 9421 9433 9560 9600 9830 9910 9920). The content of the pictures varied from human faces (18 pleasant and 15 unpleasant), animal faces (4 pleasant) and accident scenarios (5 unpleasant, e.g. a sinking ship). According to (Bradley and Lang, 2007), the selected pleasant affective pictures should induce pleasant moods in subjects and the selected unpleasant affective pictures should induce unpleasant moods. Each type consisted of 32 slides of affective pictures (about 3-minute long) presented at intervals of 6 seconds. The whole set of the affective pictures was shown to the subjects through Microsoft Office PowerPoint 2007. Each presentation began with a slide saying: “Each picture will be shown every 6 seconds. Please also pay attention to the robot head.”

56 volunteers with an average age of 24.5 participated in this experiment. These 23 male and 33 female participants had various nationalities (21 British, 15 Chinese, 9 other Asian participants, and 7 other European participants). They were recruited from the university mailing list, and they were offered the chance to win a book token worth 50-pounds. There were 14 subjects in each group, resulting in 28 subjects in the Congruent Condition (group 1 and group 4), and 28 subjects in the Conflicting Condition (group 2 and group 3). Care was taken to ensure that no subjects had participated in the previous experimental sessions.

4.4.1 Experimental Results

The responses to the questionnaire administered after viewing the robot were analysed. Table 4.9 shows subjects' perception of the facial expressions of the robot when they watched the affective pictures at the same time: they were not particularly good at reliably identifying the emotional expressions of the robot. A response was considered to be correct when the robot head was said to have shown Positive Affect in the Positive Affect Condition, or Negative Affect in the Negative Affect Condition (the neutral choice was counted as wrong in both conditions). It seems that subjects' judgements about the robot were affected by the accompanying affective pictures. When the robot's positive expressions were accompanied by "positive" pictures, they were correctly recognised as such 78.6% of the time, as compared to only 14.3% of the time when accompanied by "negative" pictures. Conversely, the robot's negative expressions were correctly recognised as such 92.9% of the time when paired with "negative" pictures, as opposed to 57.1% of the time when paired with "positive" pictures.

Table 4.9 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Image1.

%match		Positive Affect	Neutral Affect	Negative Affect	% correct
Positive Affect Condition	group 1 (congruent pictures)	78.6	7.1	14.3	78.6
	group 3 (conflicting pictures)	14.3	7.1	78.6	14.3
Negative Affect Condition	group 2 (conflicting pictures)	21.4	21.5	57.1	57.1
	group 4 (congruent pictures)	7.1	0	92.9	92.9

A Fisher's exact test in PASW 18 with two-tailed $p<0.0001$ indicated a significant association between information style (Conflicting Information or Congruent Information) and accuracy of attributing robot emotions to its facial expressions, in other words, subjects were less able to recognize the robot's emotional expressions, when the expressions conflicted with valence of the affective pictures (a lower accuracy, 10/28 (35.7% correct, 64.3% incorrect)), than when the robot behaved congruently with the affective pictures (a much higher accuracy, 24/28 (85.7% correct, 14.3% incorrect)). In addition, a Chi-square test for independence (with Yates Continuity Correction) in PASW 18 stated that $\chi 2$ (1, n=56) =12.652, two-tailed p<0.0001, correlation coefficient phi=0.512 (large effect). Consequently, H1 that subjects will recognize the sequence of robotic facial expressions better in a congruent context than in a conflicting context, was supported.

As shown in Table 4.10, subjects were mostly good at recognizing the emotional content of the pictures presented at the same time as the robot's expressions. A response was considered to be correct when a set of affective pictures was said to have been Positive in the Positive Pictures Condition, or Negative in the Negative Pictures Condition (a neutral response was counted as incorrect in either condition). It seemed that subjects' identification of the emotion conveyed by the affective pictures was not affected by whether or not the robot face displayed matching emotions.

Table 4.10 Percentage of three possible labels (items in the first row) chosen to match the displayed set of affective pictures (positive or negative) in Image1.

%match		Positive Pictures	Neutral Pictures	Negative Pictures	% correct
Positive Pictures Condition	group 1 (congruent robot)	100	0	0	100
	group 2 (conflicting robot)	96	4	0	96
Negative Pictures Condition	group 3 (conflicting robot)	0	0	100	100
	group 4 (congruent robot)	0	4	96	96

The hypothesis (H2) that when the emotional valences of the context and the robot face are conflicting, the context will have a more dominant effect over the robot face in attributing the robot's facial expressions to its emotions, was also tested. A Fisher's exact test in PASW 18 with two-tailed $p<0.0001$ indicated the affective pictures modality had a stronger perceptual effect than the Robot Expression modality when subjects were presented with conflicting information (27/28 accuracy for BBC News VS. 10/28 accuracy for Robot Emotions). In addition, a Chi-square test for independence (with Yates Continuity Correction) in PASW 18 stated that $\chi 2$ (1, n=56) =20.393, two-tailed $p<0.0001$, correlation coefficient phi=-0.641 (large effect). Consequently, H2 was supported.

4.4.2 Discussion and Conclusion

H1 and H2 were validated in this experiment. To sum up the findings in this experiment (Image1), it was observed that, like the recorded BBC News in Speech1, the selected affective pictures in this experiment, exerted a contextual effect on subjects' attribution of the robotic facial expressions to its emotions. Situation information dominance was found here for selected affective pictures as well as for the recorded BBC News.

As mentioned before, contextual effects were found with human faces and avatar faces, and these faces were often paired with voices which were used as the contexts. Contextual effects were also found with the robot face investigated here, in Speech1 and Speech2 with the particular speech audio as a surrounding context. These previous studies on human faces, avatar faces and a particular robot face have observed contextual effects when subjects were perceiving the emotional cues from two different channels, such as visual (the robot face) - audio (the News recordings) in Speech1 and Speech2. How this experiment differed from experiment Speech1 and some other studies on human faces and avatar faces, was that still images were used as a surrounding context, rather than voices. And this experiment confirmed that still images as a new form of context can also exert a contextual effect, when subjects were perceiving emotional cues from the same channel (i.e., visual channel (still images) with visual channel (robot face)).

Since a contextual effect was found for still images by adapting the same methodology used in Speech1, the next step is to explore the circumstances under which this kind of emotional context can affect people's view of the robot's expressions. Can images as a surrounding context exert a contextual effect even on a second viewing of the robot face, as the news recordings did in Speech2? A further experiment involving still images will be conducted to answer the above question. It is expected that a contextual effect will be still found on affective pictures, as it was found in Speech2 even on a second viewing of the robot face.

4.5 Experiment Image2

Context: affective pictures

Manner of Presentation: simultaneously presented with the robot's expressions

Viewing Order of the Robot Face: second viewing

Mood States of the Subjects: not measured

Subjects were again divided into four groups according to different combinations of affective pictures (positive vs. negative) and robot expressions (positive vs. negative) in the manner as described in section 4.1.5.

60 volunteers with an average age of 22.63 participated in this experiment. These 29 male and 31 female participants had various nationalities (32 British, 9 Chinese, 16 other Asian participants, and 2 other European participants).

They were recruited from leaflets distributed to the locations such as the Student Union, some student accommodations in the university, and some lecture rooms in the university. Each participant was paid £5 for each taking part, as was advertised on every leaflet. There were 15 subjects in each group, resulting in 30 subjects in the Congruent Condition (group 1 and group 4), and 30 subjects in the Conflicting Condition (group 2 and group 3). Care was taken to ensure that no subjects had participated in the previous experimental sessions.

4.5.1 Experimental Results

The responses to the questionnaire administered after viewing the robot were analyzed. Table 4.11 shows subjects' perception of the facial expressions of the robot when they watched the affective pictures at the same time: they were not particularly good at reliably identifying the emotional expressions of the robot. A response was considered to be correct when the robot head was said to have shown Positive Affect in the Positive Affect Condition, or Negative Affect in the Negative Affect Condition (the neutral choice was counted as wrong in both conditions). It seemed that subjects' judgments about the positive robot facial expressions were affected by the accompanying affective pictures. When the robot's positive expressions were accompanied by "positive" pictures, they were correctly recognized as such 93.3% of the time, as compared to only 46.7% of the time when accompanied by "negative" pictures.

Table 4.11 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Image2.

%match		Positive Affect	Neutral Affect	Negative Affect	% correct
Positive Affect Condition	group 1 (congruent pictures)	93.3	0	6.7	93.3
	Group 3 (conflicting pictures)	46.7	0	53.3	46.7
Negative Affect Condition	group 2 (conflicting pictures)	6.7	0	93.3	93.3
	group 4 (congruent pictures)	0	0	100	100

Moreover, A Fisher's exact test in PASW 18 with two-tailed p=0.12 indicated a significant association between Information Style (Conflicting Information or Congruent Information) and Accuracy of recognizing robot's emotions, in other words, accuracy of subjects' perception of robot's emotions was significantly different between when the robot behaved conflictingly with the affective pictures (a medium accuracy, 21/30 (70.0% correct, 30% incorrect)), and when the robot behaved congruently with the affective pictures (a relatively higher accuracy, 29/30 (96.7% correct, 3.3% incorrect)). In addition, a Chi-square test for independence (with Yates Continuity Correction) in PASW 18 stated that $\chi2$ (1, n=60) =5.880, two-tailed p=0. 015, correlation coefficient phi=-0.358 (medium effect). Consequently, H1 that subjects will recognize the sequence of robotic facial expressions better in a congruent context than in a conflicting context, was supported.

Table 4.12 shows subjects' perception of the affective pictures when they watched the robot emotions at the same time: subjects were perfectly good at identifying the emotional content of the pictures. A response was considered to be correct when a set of affective pictures was said to have been Positive in the Positive Pictures Condition, or Negative in the Negative Pictures Condition (a neutral response was counted as incorrect in either condition). It seemed that subjects' identification of the emotion conveyed by the affective pictures was not affected by whether or not the robot face displayed matching emotions.

Table 4.12 Percentage of three possible labels (items in the first row) chosen to match the displayed set of affective pictures (positive or negative) in Image2.

%match		Positive Pictures	Neutral Pictures	Negative Pictures	% correct
Positive Pictures Condition	group 1 (congruent robot)	100	0	0	100
	group 2 (conflicting robot)	100	4	0	96
Negative Pictures Condition	group 3 (conflicting robot)	0	0	100	100
	group 4 (congruent robot)	0	4	100	100

The hypothesis H2 that when the emotional valences of the context and the robot face are conflicting, the context will have a more dominant effect over the robot face in attributing the robot's facial expressions to its emotions, was also tested. A Fisher's exact test in PASW 18 with two-tailed p=0.002 indicated the affective pictures had a more dominant effect on the judgments about the emotions of the robot than the robot emotions themselves (30/30 accuracy for affective pictures VS. 21/30 accuracy for Robot Emotions). In addition, a Chi-square test for independence (with Yates Continuity Correction) in PASW 18 stated that $\chi 2$ (1, n=60) =8.366, two-tailed p=0. 004 (although the data violated the assumption of equal variances), correlation coefficient phi=0.42 (medium effect). Consequently, H2 was supported.

4.5.2 Discussion and Conclusion

H1 and H2 were validated, showing that the selected affective pictures also produced a contextual effect despite the fact that the robot face and its expressions were being viewed for the second time, just as the News recordings did in Speech2. And similarly to the results in Image1, situation information dominance was also found for the selected affective pictures in this experiment by using a procedure and materials that were identical to those used in Image1. As in Image1, still images in this experiment were presented in a way that subjects would not think of the robot was capable of “seeing” the images.

This experiment reconfirms that the selected affective pictures as a new form of context, can exert a contextual effect just as the speech audio did. And such still images were found to be more dominant over the robot face in attributing its facial expressions to its emotions, which indicated that a second viewing of the robot face did not weaken a contextual effect significantly.

So far, efforts have been made to explore the relative effects of two different kinds of surrounding context. And for each form of context, in order to see the circumstances under which such contextual effects are likely to occur, further investigation of whether subjects have seen the robot face before makes a difference was carried out. However, it remains uncertain whether the manner in which the context and the robot head are presented (simultaneously or separately) makes a difference. As mentioned before, if the context under consideration is presented before the synthetic robot emotions are viewed, instead of simultaneously, the context may not exert the same effect as it does when the head and context are viewed simultaneously. This inspires a method of examination, which is separating the context and the robot’s behavior in time, so that they do not occur simultaneously. In the next experiment, this question will be probed further, through the use of the same affective pictures, but presenting them ***before*** the presentation of the robot’s sequence of emotional expressions.

4.6 Experiment Image3

Context: affective pictures

Manner of Presentation: presented before the presentation of the robot's expressions

Viewing Order of the Robot Face: first viewing

Mood States of the Subjects: not measured

Subjects were again divided into four groups according to different combinations of affective pictures (positive vs. negative) and robot expressions (positive vs. negative) in the manner as described in section 4.1.5.

30 volunteers with an average age of 28 participated in this experiment. These 14 male and 16 female participants had various nationalities (8 British, 6 Chinese, 7 other Asian participants, and 4 other European participants). They were recruited from the university mailing list, and they were offered the chance to win a 50-pound worth of book token. There were either 7 or 8 subjects in each group, resulting in 15 subjects in the Congruent Condition (8 subjects in group 1 and 7 subjects in group 4), and 15 subjects in the Conflicting Condition (7 subjects in group 2 and 8 subjects in group 3). Care was taken to ensure that no subjects had participated in the previous experimental sessions.

4.6.1 Experimental Results

The responses to the questionnaire administered after viewing the robot were analyzed. Table 4.13 shows subjects' perception of the facial expressions of the robot after they watched the affective pictures: they were good at reliably identifying the emotional expressions of the robot except for group 2. A response was considered to be correct when the robot head was said to have shown Positive Affect in the Positive Affect Condition, or Negative Affect in the Negative Affect Condition (a neutral response was counted as incorrect in either condition). It seemed that subjects' judgments about the robot were affected by the preceding affective pictures. When the robot's positive expressions were preceded by "positive" pictures, they were correctly recognized as such 87.5% of the time, as compared to 75% of the time when preceded by "negative" pictures. Conversely, the robot's negative expressions were correctly recognized as such 100% of the time when paired with "negative" pictures, as opposed to 57.1% of the time when paired with "positive" pictures.

However, statistical testing found no evidence of a significant effect of the affective pictures context on subjects' identification of the robot's emotional expressions when the pictures preceded them. This is in contrast to the contextual effect obtained in Experiment Image1 when affective pictures were presented at the same time as the robot's expressions. A Fisher's exact test in PASW 18 with two-tailed p=0. 1686 indicated no significant association between information style (Conflicting Information or Congruent Information) and accuracy of attribution of robot emotions to its facial expressions. In other words, accuracy of subjects' perception of robot's emotions was not significantly different between when the robot behaved conflictingly with the affective pictures (a medium accuracy, 10/15 (66.7% correct, 33.3% incorrect)), and when the robot behaved congruently with the affective pictures (a relatively higher accuracy, 14/15 (93.3% correct, 6.7% incorrect)). Consequently, H1 was not supported in this case. In other words, the preceding context, the affective pictures in this case, did not color subjects' perception of the synthetic robot emotions, which leads to a conclusion that an emotional context will only affect the attribution of robot emotions to its facial expressions when that context is presented simultaneously and not when it precedes the robot's sequence of facial expressions.

Table 4.13 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Image3.

%match		Positive Affect	Neutral Affect	Negative Affect	% correct
Positive Affect Condition	group 1 (congruent pictures)	87.5	12.5	0	87.5
	group 3 (conflicting pictures)	75	12.5	12.5	75
Negative Affect Condition	group 2 (conflicting pictures)	0	42.9	57.1	57.1
	group 4 (congruent pictures)	0	0	100	100

As shown in Table 4.14, subjects' were mostly good at identifying the emotional content of the pictures. A response was considered to be correct when a set of affective pictures was said to have been Positive in the Positive Pictures Condition, or Negative in the Negative Pictures Condition (a neutral response was counted as incorrect in either condition). And it also seems that subjects' identification of the emotion conveyed by the affective pictures was not affected by whether or not the robot face displayed matching emotions.

Table 4.14 Percentage of three possible labels (items in the first row) chosen to match the displayed set of affective pictures (positive or negative) in Image3.

%match		Positive Pictures	Neutral Pictures	Negative Pictures	% correct
Positive Pictures Condition	group 1 (congruent robot)	87.5	12.5	0	87.5
	group 2 (conflicting robot)	71.4	28.6	0	71.4
Negative Pictures Condition	group 3 (conflicting robot)	0	0	100	100
	group 4 (congruent robot)	0	0	100	100

4.6.2 Discussion and Conclusion

H1 was not validated, and there was no need to test H2 in this experiment. This indicates that images presented as a preceding context did not exert a contextual effect, which suggests in turn that timing plays an important role in human's perception of synthetic robot facial expressions. Contextual effects were found in the earlier experiments for speech audio and still images under conditions of simultaneous presentation of the context and the robot face, on either a first viewing or a second viewing. It can be speculated that a context will affect subjects' recognition of the robot expressions only when the context is presented simultaneously with the robot's sequence of facial expressions but not when it precedes them. In other words, the manner in which the context and the robot head are presented (simultaneously or separately) makes a difference to the presence of a situational influence.

Further experiments with different kinds of context will be conducted to find out if it is always the case that an absence of a contextual effect is observed when the context, and robot expressions are presented separately. However, before doing so, some further experiments involving the simultaneous presentation of a new kind of context and the robot face will be conducted firstly. This is because the main purpose of this thesis is to find out what kinds of context can be shown to exert a contextual effect on the recognition of the robotic facial expressions. Therefore a new form of context, classical music pieces, will be used in the next experiment, whilst employing the same experimental methodology as used in Speech1 and Image1. It is expected that music presented as a surrounding context will have a situational influence on the recognition of the same robotic facial expressions on a first viewing of the robot face.

4.7 Experiment Music1

Context: classical music

Manner of Presentation: simultaneously presented with the robot's expressions

Viewing Order of the Robot Face: first viewing

Mood States of the Subjects: not measured

Subjects were divided into four groups according to different combinations of classical music (happy vs. sad) and robot expressions (positive vs. negative) in the manner as described in section 4.1.5.

The 3 minute long classical music piece used in the experiment was chosen from a list in provided in (Eich et al, 2007). Subjects listened to two types of classical music, one was happy music (either Beethoven's Symphony #9: Presto (1min) plus Tchaikovsky's 1812 Overture (excerpt) (2mins) or Tchaikovsky's The Nutcracker: Dance of the Flutes (2mins) plus Tchaikovsky's The Nutcracker: Trepak (1min)), and the other was sad music (either Albinoni's Adagio in G Minor (3min) or Stravinsky's Firebird: Lullaby (excerpt) (3mins)). According to (Eich et al, 2007), the selected happy classical music will induce happy moods in subjects and the selected sad classical music will induce sad moods.

60 volunteers with an average age of 22.63 participated in this experiment. These 29 male and 31 female participants had various nationalities (32 British, 9 Chinese, 16 other Asian participants, and 2 other European participants). They were recruited from leaflets distributed to the locations such as the Student Union, some student accommodations in the university, and some lecture rooms in the university. Each participant was paid £5 for each taking part, as was advertised on every leaflet. There were 15 subjects in each group, resulting in 30 subjects in the Congruent Condition (group 1 and group 4), and 30 subjects in the Conflicting Condition (group 2 and group 3). Care was taken to ensure that no subjects had participated in the previous experimental sessions.

4.7.1 Experimental Results

The responses to the questionnaire administered after viewing the robot were analyzed. Table 4.15 demonstrates subjects' perception of the facial expressions of the robot when they listened to the classical music simultaneously: they were good at reliably identifying the emotional expressions of the robot except for group 2 and group 3. A response was considered to be correct when the robot head was said to have shown Positive Affect in the Positive Affect Condition, or Negative Affect in the Negative Affect Condition (a neutral response was counted as incorrect in either condition). It seems that subjects' judgments about the robot were affected by the accompanying classical music. When the robot's positive expressions were accompanied by "happy" music, they were correctly recognized as such 100% of the time, as compared to only 60% of the time when accompanied by "sad" music. Conversely, the robot's negative expressions were correctly recognized as such 86.7% of the time when paired with "sad" music, as opposed to 40% of the time when paired with "happy" music.

Moreover, statistical analysis showed these differences were significant. A Fisher's exact test in PASW 18 with two-tailed $p<0.0001$ indicated a significant association between Information Style (Conflicting Information or Congruent Information) and Accuracy of recognizing robot's emotions. In other words, the accuracy of subjects' perception of robot's emotions was significantly different depending on whether the robot behaved conflictingly with the classical music (a medium accuracy, 15/30 (50% correct, 50% incorrect)), or whether when the robot behaved congruently with the classical music (a relatively higher accuracy, 28/30 (93.3% correct, 6.7 % incorrect)). In addition, a Chi-square test for independence (with Yates Continuity Correction) in PASW 18 stated that $\chi 2$ (1, n=60) =11.819, two-tailed p=0.001, correlation coefficient phi=-0.481 (medium effect). Consequently, H1 that subjects will recognize the sequence of robotic facial expressions better in a congruent context than in a conflicting context, was supported in this case.

Table 4.15 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Music1.

%match		Positive Affect	Neutral Affect	Negative Affect	% correct
Positive Affect Condition	group 1 (congruent music)	100	0	0	100
	group 3 (conflicting music)	60	6.7	33.3	60
Negative Affect Condition	group 2 (conflicting music)	53.3	6.7	40	40
	group 4 (congruent music)	6.7	6.6	86.7	86.7

Table 4.16 illustrates subjects' perception of the classical music when they watched the robot emotions simultaneously: they were mostly good at identifying the emotional content of the happy classical music, but less good at identifying the emotional content of the sad classical music. A response was considered to be correct when a piece of classical music was said to have been happy in the Happy Music Condition, or sad in the Sad Music Condition (a neutral response was counted as incorrect in either condition). It seems that in this experiment, subjects' identification of the emotion conveyed by the classical music was affected by whether or not the robot face displayed matching emotions. The Happy music was correctly recognized as such 100% of the time when it was accompanied by a "happy" robot, and only 73% of the time when accompanied by a "sad" robot. The Sad music was recognized as such 93% of the time when accompanied by a "sad" robot, and only 53% of the time when accompanied by a "happy" robot.

Interestingly, although it was not one of the hypotheses, it was found that the synthetic robot emotions seemed to affect subjects' perception of the classical music. A Fisher's exact test in PASW 18 with two-tailed p=0.02 indicated a significant association between Information Style (Conflicting Information or Congruent Information) and Accuracy of recognizing the classical music. In other words, there was a significant difference in the accuracy of the subjects' perception of the classical music depending on whether the classical music and the robot's expressions showed conflicting emotions (a medium accuracy, 19/30 (63.3% correct, 36.7% incorrect)), or congruent emotions (a relatively higher accuracy, 29/30 (96.7% correct, 3.3 % incorrect)). In addition, a Chi-square test for independence (with Yates Continuity Correction) in PASW 18 stated that $\chi 2$ (1, n=60) =8.438, two-tailed p=0.004, correlation coefficient phi=-0.417 (medium effect).

Further statistical tests were applied to see whether H2 that when the emotional valences of the context and the robot face are conflicting, the context will have a more dominant effect over the robot face in attributing the robot's facial expressions to its emotions, was supported. A Fisher's exact test in PASW 18 with two-tailed p=0.435 indicated the classical music did not have a more dominant effect on the judgments about the emotions of the robot than the robot emotions themselves (19/30 accuracy for classical music VS. 15/30 accuracy for Robot Emotions). Consequently, H2 was not supported in this case.

Table 4.16 Percentage of three possible labels (items in the first row) chosen to match the played piece of classical music (happy or sad) in Music1.

%match		Happy Music	Neutral Music	Sad Music	% correct
Happy Music Condition	group 1 (congruent robot)	100	0	0	100
	group 2 (conflicting robot)	73.3	26.7	0	73.3
Sad Music Condition	group 3 (conflicting robot)	26.7	20	53.3	53.3
	group 4 (congruent robot)	0	6.7	0	93.3

4.7.2 Discussion and Conclusion

This experiment has examined how a surrounding context (congruent or incongruent classical music) influenced users' perception of a robot's simulated emotional expressions. Hypothesis H1, that subjects will recognize the sequence of robotic facial expressions better in a congruent context than in a conflicting context, was supported in this case. It seems that music as a new form of surrounding context can be shown to exert a contextual effect on the first viewing of the robot face, as well as recorded speech in Speech1, and affective pictures in Image1. However, the second hypothesis H2, that when the emotional valences of the context and the robot face are conflicting, the context will have a more dominant effect over the robot face in attributing the robot's facial expressions to its emotions, was not validated. It was found that the contextual influence was bidirectional: the music influenced judgments of the robot's emotional expressions, and the robot's expressions affected people's judgments of the emotional valence of the music.

The results reported here cohere with the limited situational dominance account (Carroll and Russell, 1996), although that account was proposed in the context of the recognition and interpretation of human facial expressions. The limited situational dominance account views the facial expression itself as only one element in the interpretation of emotional expressions – the others being the surrounding situation, and also the current state of the observer. The present study extends the knowledge about the kinds of context that affect the recognition of robot emotional expressions. It seems that emotionally valenced music can be shown to affect such recognition, as well as recorded speech with an emotional content, and affective pictures. At the same time, the present results indicate that music has a different influence than the surrounding situations previously investigated. The results obtained here indicated a bidirectional effect of a musical context, which is consistent with earlier studies (De Gelder and Vroomen, 2000; Hong et al. 2002; Creed and Beale, 2008) with avatars or human beings. This is in contrast to the earlier experiments in which other emotional contexts were found to have a more dominant effect over the robot face – subjects were more affected by the emotional content of BBC news, and by the content of affective pictures in their judgments of the robot's emotions than they were by the content of the expressions themselves. Some other researchers also found a stronger effect of surrounding context on the recognition of avatar and human faces (Carroll and Russell, 1996; Noël et al. 2009; Mower et al. 2008; Mower et al. 2009). It seems that researchers do not yet have a full understanding of the factors that determine the relative effects of different forms of context.

In summary, the results reported here show that emotionally valenced music, in particular, can exert a contextual effect on the recognition of the emotional expressions of a robot (although its effects are bidirectional in contrast to previously investigated pictorial and spoken contexts). In earlier experiments with speech audio and still images it was found that contextual effects still occurred, and were not weakened, even if subjects had already seen the robot producing a sequence of expressions before. The next experiment investigates whether music used as a surrounding context will still exert a contextual effect, even on a second viewing of the robot face producing a sequence of expressions.

4.8 Experiment Music2

Context: classical music

Manner of Presentation: simultaneously presented with the robot's expressions

Viewing Order of the Robot Face: second viewing

Mood States of the Subjects: measured

Subjects were again divided into four groups according to different combinations of classical music (happy vs. sad) and robot expressions (positive vs. negative) in the manner as described in section 4.1.5.

60 volunteers with an average age of 21.9 participated in this experiment. These 33 male and 27 female participants had various nationalities (23 British, 22 Chinese, 11 other Asian participants, and 3 other European participants). They were recruited from leaflets distributed to the locations such as the Student Union, some student accommodations in the university, and some lecture rooms in the university. Each participant was paid £5 for each taking part, as was advertised on every leaflet. There were 15 subjects in each group, resulting in 30 subjects in the Congruent Condition (group 1 and group 4), and 30 subjects in the Conflicting Condition (group 2 and group 3). Care was taken to ensure that no subjects had participated in the previous experimental sessions.

Subjects' mood states were measured twice - they filled in two copies of the Affect Grid before the short questionnaire, once before and once after the presentation of the emotional music.

4.8.1 Experimental Results

The responses to the questionnaire administered after viewing the robot were analyzed. Table 4.17 illustrates subjects' perception of the facial expressions of the robot when they listened to the classical music simultaneously: they were good at reliably identifying the emotional expressions of the robot except for group 2. A response was considered to be correct when the robot head was said to have shown Positive Affect in the Positive Affect Condition, or Negative Affect in the Negative Affect Condition (a neutral response was counted as incorrect in either condition). It seems that subjects' judgments about the robot were affected by the accompanying classical music. When the robot's positive expressions were accompanied by "happy" music, they were correctly recognized as such 100% of the time, as compared to only 86.7% of the time when accompanied by “sad” music. Conversely, the robot’s negative expressions were correctly recognized as such 100% of the time when paired with “sad” music, as opposed to 60% of the time when paired with “happy” music.

Moreover, statistical analysis showed these differences were significant. A Fisher’s exact test in PASW 18 with two-tailed p=0.005 indicated a significant association between information style (congruent or conflicting) and accuracy of attributing robot emotions to its facial expressions. In other words, the accuracy of subjects’ perception of robot’s emotions was significantly different depending on whether the robot behaved congruently with the classical music (a perfect accuracy, 30/30 (100% correct, 0% incorrect)), or whether the robot behaved conflictingly with the classical music (a high accuracy, 22/30 (73.3% correct, 26.7% incorrect)). In addition, a Chi-square test for independence (with Yates Continuity Correction) in PASW 18 stated that $\chi2$ (1, n=60) =7.076, two-tailed p=0. 008 (although the data violated the assumption of equal variances), correlation coefficient phi=0.392 (medium effect). Consequently, H1 that subjects will recognize the sequence of robotic facial expressions better in a congruent context than in a conflicting context, was supported in this case.

Table 4.17 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Music2.

%match		Positive Affect	Neutral Affect	Negative Affect	% correct
Positive Affect Condition	group 1 (congruent music)	100	0	0	100
	group 3 (conflicting music)	86.7	6.7	6.6	86.7
Negative Affect Condition	group 2 (conflicting music)	26.7	13.3	60	60
	group 4 (congruent music)	0	0	100	100

Table 4.18 illustrates subjects' perception of the classical music when they watched the robot emotions simultaneously: they were mostly good at identifying the emotional content of the happy classical music, but less good at identifying the emotional content of the sad classical music. A response was considered to be correct when a piece of classical music was said to have been happy in the Happy Music Condition, or sad in the Sad Music Condition (a neutral response was counted as incorrect in either condition). It seems that subjects' identification of the emotion conveyed by the sad classical music was affected by whether or not the robot face displayed matching emotions.

However, in this study a Fisher's exact test in PASW 18 showed that whether or not the robot face displayed matching emotions did not affect subjects' perception of the classical music significantly. As the accuracy of subjects' perception of classical music was not significantly different depending on whether it behaved congruently with the robot face (a high accuracy, 27/30 (90% correct, 10% incorrect)), or whether it behaved conflictingly with the robot face (a high accuracy, 25/30 (83.3% correct, 16.7% incorrect)) with two-tailed p=0.706.

It is also interesting to see whether H2 that when the emotional valences of the context and the robot face are conflicting, the context will have a more dominant effect over the robot face in attributing the robot's facial expressions to its emotions, can be supported. Subjects watched the synthetic robot emotions while listening to the classical music simultaneously, having seen the opposite valence of the robot emotions before the presentation of the classical music, in this experiment. A Fisher's exact test in PASW 18 with two-tailed p=0.532 indicated the classical music did not have a more dominant effect on the judgments about the emotions of the robot than the robot emotions themselves (25/30 accuracy for classical music VS. 22/30 accuracy for Robot Emotions). Consequently, H2 was not supported in this case.

Table 4.18 Percentage of three possible labels (items in the first row) chosen to match the played piece of classical music (happy or sad) in Music2.

%match		Happy Music	Neutral Music	Sad Music	% correct
Happy Music Condition	group 1 (congruent robot)	86.7	13.3	0	86.7
	group 2 (conflicting robot)	100	0	0	100
Sad Music Condition	group 3 (conflicting robot)	20	13.3	66.7	66.7
	group 4 (congruent robot)	0	6.7	93.3	93.3

A two-way between-groups ANOVA in PASW 18 was conducted to explore the impact of music and robotic facial expressions on subjects' mood states, as measured by the Affect Grid (see the process of measurement in Appendix Ⅵ: The Procedures of Experiments Video1 and Music2 Conducted in Pairs). First of all, the data did not violate the homogeneity of variances assumption, as a significant result (Sig. value equals to 0.311) suggests that the variance of the dependent variable, i.e., the mood states, across the groups is equal. The interaction effect between music and robotic facial expressions groups was not statistically significant, F (1, 56) =.265, p = 0.609. There was a statistically significant main effect for music (Happy Music: M =3.3667, SD = 2.39947, and Sad Music: M =-1.8333, SD= 2.82944), F (1, 56) = 64.454, p < 0.0001; and the effect size was large (partial eta squared =.535). The main effect for robotic facial expressions (Robot's Positive Affect: M =1.6333, SD =3.62447, and Robot's Negative Affect: M =-.1000, SD= 3.61367), F (1, 26) = 7.162, p=0.010, did reach statistical significance and the effect size was small (partial eta squared =0.113). Cohen's eta squared represents the proportion of variance in the dependent variable that is explained by the independent (group) variable (for 0.01=small effect, 0.06=moderate effect, and 0.14=large effect). In other words, the sequences of robotic facial expressions did not have an effect on subjects' moods, while the musical pieces were found to have a strong effect on the subjects' moods.

4.8.2 Discussion and Conclusion

H1 was validated but H2 was not validated; findings that resemble those obtained in the Music1 experiment. Similar to the findings in Speech2 and Image2, even on a second presentation of the robot's emotional expressions, the classical music context was able to color subjects' perception of the synthetic robot facial expressions that occurred with it. Like the results in Music1, no situational dominance was found for the classical music, but the results of this experiment differed from Music1 in that the contextual influence reported here was not bidirectional: the robot's expressions did not affect people's judgments of the emotional valence of the music. There is no obvious reason for this difference, since the same musical pieces were used in both. The only substantive difference between this experiment and Music 1 was that the robot head was being viewed for the second time, but there is no reason to think that this would affect the extent of the influence of the music on the subjects' moods. The results obtained here nonetheless re-confirmed the conclusion in Music1 that music has a different situational influence than the surrounding contexts previously investigated, since no evidence of a dominant effect of context over the robot face was found.

As mentioned earlier, the limited situational dominance account views the facial expression itself as only one element in the interpretation of emotional expressions – others being the surrounding situation, and also the current state of the observer. There was evidence in this experiment, provided by two-way ANOVA in PASW 18 that the happy music made people feel happier than the sad music, which indicated that subjects' mood states were colored by the musical context. It can be speculated that this mood effect may have contributed to the contextual effect observed.

As discussed in Image3, future experiments with a different kind of context should be conducted to find out whether the separate presentation of a context always results in an absence of a contextual effect on the recognition of the robot's emotions. Therefore the next experiment involves the separate presentation of a musical context, where the music precedes the presentation of the robot's expressions. It is expected that no contextual effect will be found for the preceding music.

4.9 Experiment Music3

Context: classical music

Manner of Presentation: presented before the presentation of the robot's expressions

Viewing Order of the Robot Face: second viewing

Mood States of the Subjects: measured

Subjects were again divided into four groups according to different combinations of classical music (happy vs. sad) and robot expressions (positive vs. negative) in the manner as described in section 4.1.5.

The 20 minute pieces of classical music used in this experiment were based on selections of happy and sad music listed in (Eich et al. 2007). Two types of the classical music were again used in this study, one happy and one sad. The happy music was either set 1 (Mozart's Eine Kleine Nachtmusik: Allegro, Eine Kleine Nachtmusik: Rondo and Serenade #9: Finale, Bach's Brandenburg Concerto #3: Allegro, Tchaikovsky's The Nutcracker: Waltz of the Flowers, and The Nutcracker: Trepak) or set 2 (Beethoven's Symphony #9: Presto, Vivaldi's Four Seasons: Spring I Allegro, Four Seasons: Spring III Allegro, Four Seasons: Autumn I Allegro, and Four Seasons: Autumn III Allegro, Tchaikovsky's The Nutcracker: Dance of the Flutes and 1812 Overture (excerpt)). And the sad music was either set 1 (Albinoni's Adagio in G Minor, Sibelius's Violin Concerto: Adagio di Molto, Lehar's Vilja-Lied, Schumman's Traummeri, and Dvorak's Symphony #9: Largo), or set 2 (Grieg's Peer Gynt: The death of Ase, Tchaikovsky's Swan Lake: Dances des Cynges, Chopin's Prelude #4 in E Minor, Vivaldi's Four Seasons: Autumn Adagio and Stravinsky's Firebird: Lullaby (excerpt)). For each type of classical music, subjects were randomly assigned to listen to either set 1 or set 2. According to (Eich et al, 2007), the selected happy classical music will induce happy moods in subjects and the selected sad classical music will induce sad moods.

56 volunteers with an average age of 24.5 participated in this experiment. These 23 male and 33 female participants had various nationalities (21 British, 15 Chinese, 9 other Asian participants, and 7 other European participants). They were recruited from the university mailing list, and they were offered the chance to win a 50-pound worth of book token. There were 14 subjects in each group, resulting in 28 subjects in the Congruent Condition (group 1 and group 4), and 28 subjects in the Conflicting Condition (group 2 and group 3). Care was taken to ensure that no subjects had participated in the previous experimental sessions.

4.9.1 Experimental Results

The responses to the questionnaire administered after viewing the robot were analyzed. Table 4.19 shows subjects' perception of the facial expressions of the robot after they listened to the classical music: they were very good at identifying the emotional expressions of the robot. A response was considered to be correct when the robot head was said to have shown Positive Affect in the Positive Affect Condition, or Negative Affect in the Negative Affect Condition (a neutral response was counted as incorrect in either condition). It seemed that subjects' judgments about the robot were not affected by the preceding classical music.

Moreover, a Fisher's exact test in PASW 18 was conducted to test whether there was a significant association between information style (congruent or conflicting) and accuracy of attributing robot emotions to its facial expressions. It was found that the accuracy of subjects' perception of robot's emotions did not significantly differ depending on whether the robot behaved congruently with the preceding classical music (a very high accuracy, 27/28 (96.4% correct, 3.6% incorrect)), or whether the robot behaved conflictingly with the preceding classical music (also a very high accuracy, 25/28 (89.3% correct, 10.7% incorrect)) with two-tailed p=0.611. Consequently, H1 that subjects will recognize the sequence of robotic facial expressions better in a congruent context than in a conflicting context, was not supported in this case. On the basis of these results, it can be suggested that an emotional context will only affect the attribution of robot emotions to its facial expressions when that context is presented simultaneously and not when it precedes the robot's sequence of facial expressions.

Table 4.19 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Music3.

%match		Positive Affect	Neutral Affect	Negative Affect	% correct
Positive Affect Condition	group 1 (congruent music)	100	0	0	100
	group 3 (conflicting music)	92.9	7.1	6.6	92.9
Negative Affect Condition	group 2 (conflicting music)	7.1	7.2	85.7	85.7
	group 4 (congruent music)	0	7.1	92.9	92.9

As shown in Table 4.20, subjects were mostly good at identifying the emotional content of the happy classical music, but poor at identifying the emotional content of the sad classical music. A response was considered to be correct when a piece of classical music was said to have been happy in the Happy Music Condition, or sad in the Sad Music Condition (a neutral response was counted as incorrect in either condition). It seemed that subjects' identification of the emotion conveyed by the classical music was not affected by whether or not the robot face displayed matching emotions.

Table 4.20 Percentage of three possible labels (items in the first row) chosen to match the played piece of classical music (happy or sad) in Music3.

%match		Happy Music	Neutral Music	Sad Music	% correct
Happy Music Condition	group 1 (congruent robot)	92.9	7.1	0	92.9
	group 2 (conflicting robot)	85.7	7.1	7.1	85.7
Sad Music Condition	group 3 (conflicting robot)	28.6	21.4	50	50
	group 4 (congruent robot)	28.6	21.4	50	50

An independent-samples T-test in PASW 18 was conducted to compare the unpleasant-pleasant scores on the Affect Grid (see the process of measurement in Appendix Ⅳ: The Procedures of Experiments Image1 and Music3 Conducted in Pairs) for participants who listened to the happy classical music and participants who listened to sad classical music. There was a significant difference in scores for participants who listened to happy classical music (Mean M=3.3929, Standard Deviation SD= 2.34662), and participants who listened to sad classical music (M=-.5357, SD= 2.33305), with t(54)= 6.282, $p<0.0001$ (two-tailed). The magnitude of the difference in means (means difference=3.92857, 95% CI: 2.67482 to 5.18232) was very large (eta squared=0.42, for 0.01=small effect, 0.06=moderate effect, and 0.14=large effect). It seems that the happy music made people feel happier than the sad music.

4.9.2 Discussion and Conclusion

As expected, H1 was not validated, showing that music as a preceding context did not exert a contextual effect, and there was no need to test H2 in this experiment. These findings were similar to those obtained in Image3 when affective pictures were presented as a preceding context, before subjects viewed the robot's sequence of expressions. The results of this experiment provide further confirmation for the speculation made in the discussion of the Image3 experiment that a contextual effect will be only found under conditions of simultaneous presentation but not separate presentation. In this experiment the musical pieces were of a longer period - 20mins with similar musical content, while the musical pieces in Music1 and Music2 were about 3mins. The longer pieces could be expected to induce stronger moods in the subjects. And the T-test in PASW 18 showed that 20mins long musical pieces did color subjects' mood states, as did the 3mins long musical pieces in Music1 and Music2. It was argued in Music2 that a mood effect may have contributed to the contextual effect found there. However, a mood effect cannot be considered to be a sufficient factor for a contextual effect, since even though a mood effect of the music was found here, no contextual effect was obtained in this experiment.

Notice that this experiment involved a second viewing of the robot face. However, any argument that a contrast effect may have weakened the contextual effect can be viewed as being unconvincing in the light of the evidence obtained in the Speech2, Image2 and Music2 experiments that context effects can be obtained on a second viewing.

In the next experiments, the effects of videos as a further form of context will be investigated. As argued in section 4.3.2, it can be speculated that combining emotional cues from multi-modality sources (e.g., robot face and video) would be more difficult than from single-modality (e.g., robot face and still image, or robot face and music), and that this might reduce the contextual effect. In other words, the modality of the media may be important for obtaining such a contextual effect. It is also worth pointing out that video is relatively higher in intensity, complexity and attention capture. As Rottenberg et al. (2007) mentioned, there are seven key dimensions that are salient to the selection and the use of emotional materials. These key dimensions are intensity, complexity, demand characteristics, standardization, temporal resolution, ecological validity, and attention capture respectively. For instance, emotional film clips as an emotion elicitation procedure, are high in complexity and standardization, relatively high in intensity and attentional capture, low in demand characteristics and temporal resolution, and normal in ecological validity (Rottenberg et al. 2007). Other essential emotion elicitation procedures, such as the international affective picture system (IAPS; Bradley and Lang, 2007) and the MCI technique (combining music with thought to change mood, Eich et al, 2007), are high in standardization and relatively low in attentional capture.

So far, contextual effects have been found only when subjects were supposed to combine emotional cues from a single-modality (i.e., robot face and speech audio, robot face and image, and robot face and music). The next experiments will explore whether a contextual effect can be found when subjects are required to combine emotional cues from multi-modality (e.g., robot face and video). Further experiments with a new form of context, namely the film clips, will be conducted next

4.10 Experiment Video1

Context: film clips

Manner of Presentation: simultaneously presented with the robot’s expressions

Viewing Order of the Robot Face: first viewing

Mood States of the Subjects: not measured

Subjects were divided into four groups according to different combinations of film clip (amused/happy vs. sad) and robot expressions (positive vs. negative) in the manner as described in section 4.1.5.

The film clips used in this part was chosen from a list provided in (Rottenberg et al. 2007). Two types of film clips were used in this study, one was an amusing film clip (a fake orgasm scene in a restaurant (2mins and 53seconds), from the film “When Harry Met Sally”), and the other one was a sad film clip (a death scene (2mins and 45seconds) from the film “The Champ”). According to (Rottenberg et al. 2007), the selected amusing film clip will induce happy moods in subjects and the selected sad film clip will induce sad moods.

60 volunteers with an average age of 21.9 participated in this experiment. These 33 male and 27 female participants had various nationalities (23 British, 22 Chinese, 11 other Asian participants, and 3 other European participants). They were recruited from leaflets distributed to the locations such as the Student Union, some student accommodations in the university, and some lecture rooms in the university. Each participant was paid £5 for each taking part, as was advertised on every leaflet. There were 15 subjects in each group, resulting in 30 subjects in the Congruent Condition (group 1 and group 4), and 30 subjects in the Conflicting Condition (group 2 and group 3). Care was taken to ensure that no subjects had participated in the previous experimental sessions.

4.10.1 Experimental Results

The responses to the questionnaire administered after viewing the robot were analyzed. Table 4.21 illustrates subjects’ perception of the facial expressions of the robot when they watched the film clips simultaneously: they were not very good at reliably identifying the positive sequence of facial expressions. A response was considered to be correct when the robot head was said to have shown Positive Affect in the Positive Affect Condition, or Negative Affect in the Negative Affect Condition (a neutral response was counted as incorrect in either condition). The results in the table suggest that there may be an effect of the context on the subjects' judgments; they seem to be more able to correctly recognize the positive affect in the congruent condition (66.7%) than in the incongruent condition (40%), and they were better recognizing negative affect when the robot behaved congruently with the emotional valence of the film clip (93.3%) than it did not (86.7%).

However, statistical analysis showed these differences were not significant. A Fisher's exact test in PASW 18 with two-tailed p=0.252 indicated no significant association between Information Style (Conflicting Information or Congruent Information) and Accuracy of recognizing robot's emotions. In other words, the accuracy of subjects' perception of robot's emotions was not significantly different between when the robot behaved conflictingly with the film clip (a medium accuracy, 19/30 (63.3% correct, 36.7% incorrect)), and when the robot behaved congruently with the film clip (a relatively higher accuracy, 24/30 (80.0% correct, 20 % incorrect)). Consequently, H1 that subjects will recognize the sequence of robotic facial expressions better in a congruent context than in a conflicting context, was not supported in this case.

Table 4.21 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Video1.

%match		Positive Affect	Neutral Affect	Negative Affect	% correct
Positive Affect Condition	group 1 (congruent film clip)	66.7	13.3	20	66.7
	group 3 (conflicting film clip)	40	20	40	40
Negative Affect Condition	group 2 (conflicting film clip)	13.3	0	86.7	86.7
	group 4 (congruent film clip)	0	6.7	93.3	93.3

Table 4.22 illustrates subjects' perception of the film clips when they watched the robot emotions simultaneously: they were very good at identifying the emotional content of the sad film clips but slightly less good at identifying that of the amused (happy) film clips. A response was considered to be correct when a film clip was said to have been amused (happy) in the Amused (happy) Film Clip Condition, or sad in the Sad Film Clip Condition (a neutral response was counted as incorrect in either condition). There is no suggestion here that the subjects' identification of the emotion conveyed by the film clips was affected by whether or not the robot face displayed matching emotions.

Table 4.22 Percentage of three possible labels (items in the first row) chosen to match the displayed film clip (amused (happy) or sad) in Video1.

%match		Amused (Happy) Film Clip	Neutral Film Clip	Sad Film Clip	% correct
Amused (Happy) Film Clip Condition	group 1 (congruent robot)	93.3	6.7	0	93.3
	group 2 (conflicting robot)	93.3	6.7	0	93.3
Sad Film Clip Condition	group 3 (conflicting robot)	0	0	100	100
	group 4 (congruent robot)	0	0	100	100

4.10.2 Discussion and Conclusion

H1 was not validated, and there was no need to test H2 in this experiment, which suggests that the particular pair of film clips used here does not exert a contextual effect even under conditions of simultaneous presentation. It seems that the particular pair of film clips is an exception to the other contexts that were employed in these experiments. A contextual effect has been demonstrated for recorded BBC News, affective pictures, and classical music when they were presented simultaneously with the robot‘s sequence of expressions, but no such effect was found for the film context. However, the results found here do not indicate video as a new form of surrounding context does not have the ability to exert a contextual effect under conditions of simultaneous presentation, as only a pair of film clips has been investigated. More choices of film clips are needed to be investigated in future experiments.

It seems that for some reason, yet to be determined, the film clips did not produce the contextual effects produced by the other forms of context. However, it was noticed that the subjects were less good at identifying the emotional content of the amused (happy) film clips (93.3% accuracy) than that of the sad film clips (100% accuracy), which suggests that the choice of film clips used in these experiments was not good (especially the positive one). It can be speculated that a context effect might be obtained if the positive film clip used was more strongly recognized as being positive.

Can video exert a contextual effect on a seconding viewing of the robot face? And can video exert a mood effect on subjects? Although no contextual effect was found in the present experiment, such questions will be addressed in next experiment in which the same video and the robot face will be presented simultaneously but on a second viewing. The mood states of the subjects will be also be measured in next experiment. It is expected that no contextual effect will be found when the same video is presented simultaneously with a second viewing of the robot face. However the video is expected to have an influence on the moods of the subjects.

4.11 Experiment Video2

Context: film clips

Manner of Presentation: simultaneously presented with the robot’s expressions

Viewing Order of the Robot Face: second viewing

Mood States of the Subjects: measured

Subjects were again divided into four groups according to different combinations of film clip (amused/happy vs. sad) and robot expressions (positive vs. negative) in the manner as described in section 4.1.5.

30 volunteers with an average age of 28 participated in this experiment. These 14 male and 16 female participants had various nationalities (8 British, 6 Chinese, 7 other Asian participants, and 4 other European participants). They were recruited from the university mailing list, and they were offered the chance to win a 50-pound worth of book token. There were either 7 or 8 subjects in each group, resulting in 15 subjects in the Congruent Condition (8 subjects in group 1 and 7 subjects in group 4), and 15 subjects in the Conflicting Condition (7 subjects in group 2 and 8 subjects in group 3). Care was taken to ensure that no subjects had participated in the previous experimental sessions.

4.11.1 Experimental Results

The responses to the questionnaire administered after viewing the robot were analyzed. Table 4.23 reports subjects' perception of the facial expressions of the robot when they watched the film clips simultaneously: they were good at identifying the emotional expressions of the robot. A response was considered to be correct when the robot head was said to have shown Positive Affect in the Positive Affect Condition, or Negative Affect in the Negative Affect Condition (a neutral response was counted as incorrect in either condition). It seemed that subjects' judgments about the robot were affected by the accompanying film clips. When the robot's positive expressions were accompanied by "amused (happy)" film clips, they were correctly recognized as such 85.7% of the time, as compared to 71.4% of the time when accompanied by "sad" film clips. Conversely, the robot's negative expressions were correctly recognized as such 87.5% of the time when paired with "sad" film clips, as opposed to 75% of the time when paired with "amused (happy)" film clips.

However, when A Fisher's exact test in PASW 18 was conducted, no evidence of a significant association between information style (congruent or conflicting) and accuracy of attribution of robot emotions to its facial expressions was found. In other words, the accuracy of subjects' perception of robot's emotions was not significantly greater when the robot behaved congruently with the film clip (a high accuracy, 13/15 (86.7% correct, 13.3% incorrect)), than when the robot behaved conflictingly with the film clip (a high accuracy, 11/15 (73.3% correct, 26.7% incorrect)) with two-tailed p=0.651. Consequently, H1 that subjects will recognize the sequence of robotic facial expressions better in a congruent context than in a conflicting context, was not supported in this case.

Table 4.23 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Video2.

%match		Positive Affect	Neutral Affect	Negative Affect	% correct
Positive Affect Condition	group 1 (congruent film clip)	85.7	14.3	0	85.7
	group 3 (conflicting film clip)	71.4	0	28.6	71.4
Negative Affect Condition	group 2 (conflicting film clip)	25	0	75	75
	group 4 (congruent film clip)	0	12.5	87.5	87.5

Table 4.24 reports data on subjects' perception of the film clips when they watched the robot emotions simultaneously. It seems that they were mostly good at identifying the emotional content of the film clips. A response was

considered to be correct when a film clip was said to have been amused (happy) in the Amused (happy) Film Clip Condition, or sad in the Sad Film Clip Condition (a neutral response was counted as incorrect in either condition). It also seemed that subjects' identification of the emotion conveyed by the film clips was not affected by whether or not the robot face displayed matching emotions.

Table 4.24 Percentage of three possible labels (items in the first row) chosen to match the displayed film clip (amused (happy) or sad) in Video2.

%match		Amused (Happy) Film Clip	Neutral Film Clip	Sad Film Clip	% correct
Amused (Happy) Film Clip Condition	group 1 (congruent robot)	85.7	14.3	0	85.7
	group 2 (conflicting robot)	87.5	12.5	0	87.5
Sad Film Clip Condition	group 3 (conflicting robot)	0	0	100	100
	group 4 (congruent robot)	0	0	100	100

A two-way between-groups ANOVA in PASW 18 was conducted to explore the impact of film clips and robotic facial expressions on subjects' mood states, as measured by the Affect Grid (see the process of measurement in Appendix Ⅴ: The Procedures of Experiments Image3 and Video2 Conducted in Pairs). First of all, the data violated the homogeneity of variances assumption, as a significant result (Sig. value less than .05) which suggests that the variance of the dependent variable, i.e., the mood states, across the groups is not equal. However, an independent-samples T-test in PASW 18 conducted to compare the mood states of subjects who watched pleasant film clip and who watched unpleasant film clip, showed that the data did not violate the assumption of equal variances (Sig. value equals to 0.439). The interaction effect between film clips and robotic facial expressions groups was statistically significant, $F(1, 26) = 5.871$, $p = .023$; and the effect size was small (partial eta squared =0.184).

There was a statistically significant main effect for film clips (Amused (happy) Film Clip: M =2.0000, SD = 2.36039, and Sad Film Clip: M =-2.4000, SD= 2.74643), F (1, 26) = 29.145, *p*<0.0001; and the effect size was big (partial eta squared =0.529). The main effect for robotic facial expressions (Robot's Positive Affect: M =.5714, SD = 4.36268, and Robot's Negative Affect: M =-.8750, SD= 2.09364), F (1, 26) = 2.964, *p*=0.097, did not reach statistical significance. In a word, the sequences of robotic facial expressions did not have a significant effect on subjects' mood states, while the film clips had a strong effect on the subjects' moods, as did the musical pieces in Music2 and Music3.

4.11.2 Discussion and Conclusion

Neither H1 nor H2 was validated, with results that were similar to the findings in Video1. It seemed that as expected, on a second presentation of the robot's emotional expressions, the video context was unable to exert a contextual effect, although a contextual effect was found under the same circumstances in Speech2, Image2 and Music2. As similar to the finding in Video1, the subjects were less good at identifying the emotional content of the amused (happy) film clips with around 87% accuracy than that of the sad film clips with 100% accuracy, which again indicated that the particular positive film clip was not a good choice. The useful finding in this experiment was that the film clips were able to induce strong happy/sad moods to subjects. The two-way ANOVA in PASW 18 showed that the positive film clip made the observers happier than the negative film clip.

In earlier experiments, no contextual effect was found in a separate presentation of the images and the robot face in Image3 or in a separate presentation of the music and the robot face in Music3, even though the musical pieces had been shown to exert a mood effect on the subjects. It was evident that the film clips used here could also color subjects' mood states. The next experiment will provide further confirmation that videos do not exert a contextual effect when presented separately, on either first or second viewing. It is expected that no contextual effect will be found for a preceding video.

4.12 Experiment Video3

Context: film clips

Manner of Presentation: presented before the presentation of the robot's expressions

Viewing Order of the Robot Face: first viewing

Mood States of the Subjects: not measured

Subjects were again divided into four groups according to different combinations of film clip (amused/happy vs. sad) and robot expressions (positive vs. negative) in the manner as described in section 4.1.5.

52 volunteers with an average age of 23.23 participated in this experiment. These 31 male and 21 female participants had various nationalities (19 British, 12 Indian, 8 Chinese, 7 other Asian participants, and 3 other European participants). They were recruited from leaflets distributed to the locations such as the Student Union, some student accommodations in the university, and some lecture rooms in the university. Each participant was paid £5 for each taking part, as was advertised on every leaflet. There were 13 subjects in each group, resulting in 26 subjects in the Congruent Condition (group 1 and group 4), and 26 subjects in the Conflicting Condition (group 2 and group 3). Care was taken to ensure that no subjects had participated in the previous experimental sessions.

4.12.1 Experimental Results

The responses to the questionnaire administered after viewing the robot were analyzed. Table 4.25 illustrates subjects' perception of the facial expressions of the robot when they watched the film clips simultaneously: they were not so good at reliably identifying the positive sequence of facial expressions. A response was considered to be correct when the robot head was said to have shown Positive Affect in the Positive Affect Condition, or Negative Affect in the Negative Affect Condition (a neutral response was counted as incorrect in either condition). The results in the table suggest that there was no effect of the context on the subjects' judgments; they seem to correctly recognize the positive affect in the congruent condition (69.2%) as well as they do in the incongruent condition (69.2%), even though they were better recognizing negative affect when the robot behaved congruently with the emotional valence of the film clip (76.9%) than when they did not (46.2%).

Moreover, statistical analysis showed these differences were not significant. A Fisher's exact test in PASW 18 with two-tailed p=0.382 did not indicate a significant association between Information Style (Conflicting Information or Congruent Information) and Accuracy of recognizing robot's emotions, in other words, accuracy of subjects' perception of robot's emotions was not significantly different between when the robot behaved congruently with the film clips (a relatively higher accuracy, 19/26 (73.1% correct, 26.9% incorrect)) , and when the robot behaved conflictingly with the film clips (a medium accuracy, 15/26 (57.7% correct, 42.3% incorrect)). Consequently, H1 that subjects will recognize the sequence of robotic facial expressions better in a congruent context than in a conflicting context, was not supported in this case. Thus, the preceding context, the film clips in this case, did not color subjects' perception of the synthetic robot emotions, which further confirms the earlier conclusion that an emotional context will only affect the attribution of robot emotions to its facial expressions when that context is presented simultaneously and not when it precedes the robot's sequence of facial expressions.

Table 4.25 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative) in Video3.

%match		Positive Affect	Neutral Affect	Negative Affect	% correct
Positive Affect Condition	group 1 (congruent film clip)	69.2	23.1	7.7	69.2
	group 3 (conflicting film clip)	69.2	7.7	23.1	69.2
Negative Affect Condition	group 2 (conflicting film clip)	15.4	38.4	46.2	46.2
	group 4 (congruent film clip)	15.4	7.7	76.9	76.9

Table 4.26 shows subjects' perception of the film clips when they watched the robot emotions simultaneously: they were very good (100% recognition rate) at identifying the emotional content of the sad film clips but less good (80.8% recognition rate) at identifying that of the amused (happy) film clips. A response was considered to be correct when a film clip was said to have been amused (happy) in the Amused (happy) Film Clip Condition, or sad in the Sad Film Clip Condition (a neutral response was counted as incorrect in either condition). There is no suggestion here that the subjects' identification of the emotion conveyed by the film clips was affected by whether or not the robot face displayed matching emotions.

Table 4.26 Percentage of three possible labels (items in the first row) chosen to match the displayed film clip (amused (happy) or sad) in Video3.

%match		Amused (Happy) Film Clip	Neutral Film Clip	Sad Film Clip	% correct
Amused (Happy) Film Clip Condition	group 1 (congruent robot)	76.9	23.1	0	76.9
	group 2 (conflicting robot)	84.6	7.7	7.7	84.6
Sad Film Clip Condition	group 3 (conflicting robot)	0	0	100	100
	group 4 (congruent robot)	0	0	100	100

4.12.2 Discussion and Conclusion

As expected, H1 was not validated, and there was no need to test H2 in this experiment, the results being similar in this respect to the findings in Image3, and Music3. This experiment confirms that, as expected from previous results, the separate presentation of a film clip and the robot face did not produce a contextual effect. Again, the subjects were less good at identifying the emotional content of the same amused (happy) film clips with around 81% accuracy than that of the sad film clips with 100% accuracy, which was similar to the recognition accuracies in Video1 and Video2. All these recognition accuracies indicated that, the positive film clip was a poor choice, as opposed to the negative film clip.

4.13 Summary and Conclusion

This chapter has examined how surrounding or preceding contexts (i.e., congruent or incongruent recorded BBC News, selected affective pictures, classical music and film clips) influenced users' perception of a robot's simulated emotional expressions. An overview of the findings can be found in Table 4.27 below.

Table 4.27 Findings of the 11 experiments including the gender difference and cross-cultural difference (examined by Fisher's Exact Test in PASW 18)

Experiment (context, manner of presentation, viewing order)	H1 or H2 validated?	Mood effect of the media on the subjects?	Gender difference on the perception of the media or the robotic expressions observed?	Cross-cultural difference on perception of the media or the robotic expressions observed? (Western subjects VS. Non-Western subjects)
Speech1 (News recording, simultaneous, first viewing)	Both H1 and H2	No data collected	No evidence after testing	No evidence after testing
Speech2 (News recording, simultaneous, second viewing)	Both H1 and H2	No data collected	No evidence after testing	No evidence after testing
Image1 (affective pictures, simultaneous, first viewing)	Both H1 and H2	No data collected	No evidence after testing	No evidence after testing
Image2 (affective pictures, simultaneous, second viewing)	Both H1 and H2	No data collected	No evidence after testing	Yes, in the perception of robotic expressions with p=0.0148 (accuracy Western: 32/34 VS. Non-Western: 18/26)
Image3 (affective pictures, separate, first viewing)	Neither H1 nor H2	No data collected	No evidence after testing	No evidence after testing
Music1 (3mins musical pieces, simultaneous, first viewing)	H1 only	No data collected	No evidence after testing	No evidence after testing
Music2 (3mins musical pieces, simultaneous, second viewing)	H1 only	Yes	No evidence after testing	Yes, in the perception of music with p=0.0164 (accuracy Western: 19/26 VS. Non-Western: 33/34)
Music3 (20mins musical pieces, separate, second viewing)	Neither H1 nor H2	Yes	No evidence after testing	No evidence after testing
Video1 (film clips, simultaneous, first viewing)	Neither H1 nor H2	No data collected	Yes, on robotic expressions with p=0.0463 (accuracy: Male: 20/33 VS. Female: 23/27)	No evidence after testing
Video2 (film clips, simultaneous, second viewing)	Neither H1 nor H2	Yes	No evidence after testing	No evidence after testing
Video3 (film clips, separate, first viewing)	Neither H1 nor H2	No data collected	No evidence after testing	No evidence after testing

Table 4.27 highlights a few aspects as follows:

1. First of all, not all the surrounding contexts had contextual effects. Hypothesis H1, that subjects will recognize the sequence of robotic facial expressions better in a congruent context than in a conflicting context, was

supported in six cases. As illustrated in Table 4.27, they are cases of a simultaneous presentation of a robot face (first viewing) and the recorded BBC News (Speech1), a simultaneous presentation of a robot face (second viewing) and the recorded BBC News (Speech2), a simultaneous presentation of a robot face (first viewing) and selected affective pictures (Image1), a simultaneous presentation of a robot face (second viewing) and selected affective pictures (Image2), a simultaneous presentation of a robot face (first viewing) and 3mins long classical music (Music1), and a simultaneous presentation of a robot face (second viewing) and 3mins long classical music (Music2). It seems that even when they were used as a surrounding context, the particular pair of film clips did not exert a contextual effect, in contrast to the News recordings, affective pictures, and musical pieces all of which were observed to exert an effect. This suggests that different kinds of context may have different abilities in generating a situational influence on the recognition of robotic facial expressions.

2. A situational dominance was not found for every surrounding context having a contextual effect. Hypothesis H2 that when the emotional valences of the context and the robot face are conflicting, the context will have a more dominant effect over the robot face in attributing the robot's facial expressions to its emotions, was validated in four cases of Speech1, Image1, Image2 and Speech2. It was concluded that a dominant effect in Music1 was not observed because a bidirectional effect was found for the musical pieces and the robotic facial expressions. It can be concluded that the relative weights of different surrounding contexts and the same robot face in the emotional judgments of the robot face were different.

3. No contextual effect was found for contexts that preceded the robot facial expressions in all three cases that this occurred, namely a separate presentation of selected affective pictures and a robot face (first viewing) (Image3), a separate presentation of 20mins long classical music and a robot face (second viewing) (Music3), and a separate presentation of a film clip and a robot face (first viewing) (Video3). This reinforces a conclusion that, a

contextual effect only occurs in a simultaneous but not separate presentation of the robot face and a context. An absence of a context effect under conditions of separate presentation was still found even when the context was shown to affect the subjects' mood states (i.e., Video2 in Table 4.27). At the same time, the presence of contextual effects under conditions of simultaneous presentation was still found even if the observers had seen the opposite valence of robotic facial expressions before (i.e., Speech2 and Image3 in Table 4.27). Consequently, it can be concluded that the manner of a presentation (simultaneous or separate) of the robot face and the context was of vital importance in obtaining a contextual effect of a context on the robot face.

4. Gender or cross-cultural differences did not seem to play an important role in obtaining a contextual effect, as illustrated in Table 4.27. Only in one case (i.e., Music2) was a cross-cultural difference (western subjects VS. non-western subjects) on perception of the contexts or the robotic expressions observed, yet a contextual effect was still observed in this case. A gender difference on the perception of the media or the robotic expressions was only observed in the case of a simultaneous presentation of a film clip and a robot face (Video1), although no contextual effect was found in this case.

In summary, through the 11 experiments, efforts have been made in this chapter to realize the four objectives enumerated in Chapter 1: (1) to see whether the recognition of robot emotional expressions can be affected by context; (2) to see what kinds of context are likely to affect the recognition of robot emotional expressions; (3) to see the circumstances under which such contextual effects are likely to occur; (4) and to see whether when such effects occur, they are always dominant or whether they can sometimes be bidirectional (in the sense that the robot's emotional expressions also affect the recognition of the surrounding contexts).

Although it was not one of the main objectives in the conduct and design of these experiments, it is interesting to propose and discuss possible explanations for the presence or absence of contextual effects in these experiments. It should be noted that it will not necessarily be possible to clearly distinguish between them on the basis of the experimental findings reported here. Four possible explanations are proposed as follows:

1. One explanation is that from an emotional congruence perspective, recognizing the facial expressions of an emotional robot also involves embodiment (emotional states reactivation or reenactment), as argued earlier in Speech1. An effect of emotional congruence in perception of robot facial expressions was found when subjects were confronted with either congruent or conflicting emotional cues from the context and the robot face in six cases. As a result, the emotional congruence in this study may be similar to emotional congruence in perception of human emotions. It was therefore argued the theory of embodying emotion may also provide an explanation of emotional congruence in perception of robot facial expressions, although no empirical research has focused on the validation of embodied recognition of robot facial expressions.

2. The second explanation can be called 'the mood effect' - the context affects the emotional state of the observer, which in turn affects their perception of the robot. As illustrated in Table 4.27, effects of the contexts on subjects' moods were observed in three cases: namely, a simultaneous presentation of a robot face (second viewing) and the 3mins long classical music (Music2), a separate presentation of a robot face and the 20mins long classical music (Music3), and a separate presentation of a film clip and a robot face (Video3). Although a contextual effect together with a mood effect was found in Music2, it was argued that a mood effect was not a sufficient factor for the presence of a contextual effect, since even though subjects' mood states were shown to be affected by the preceding context, no contextual effect was found in Music3 and Video3. Since mood effects were not measured before and after the presentations of the context and the robot head in the other experiments in which contextual effects were found (e.g., Speech1 and Image1), the present set of experiments do not provide evidence about whether or not an effect on subjects' mood is necessary for a context effect to occur.

3. The third explanation is that it is possible that subjects interpreted the robot's expressions as if the robot was capable of "seeing" or "listening to" the context and that it could respond to the context accordingly. This explanation was previously proposed in the author's first conference paper and his journal paper. The manner of a presentation (simultaneous or separate) of the robot face and a context was found to be a key factor in the presence or absence of a contextual effect. A simultaneous presentation of the robot face and a context might increase the possibility that subjects interpreted the robot's expressions as if the robot was capable of "seeing" or "listening to" the context and it could respond to the context accordingly, while a separate presentation might reduce or remove the possibility that subjects interpreted the robot's expressions as if they responded to the context. However it is accepted here that this explanation does not stand well as contextual effects were still found in Image1 and Image2 even though the still images were presented in such a way that it was unlikely that the subjects would think that the robot was able to "see" the images, i.e., images were presented on a computer screen which was parallel to the robot head.

4. The last explanation from a pattern recognition perspective indicated that the contextual effect might be just a cue integration issue - subjects may have evaluated both sources together and chosen the context to reinterpret the robotic expressions. Again, this is related to the manner of a presentation (simultaneous or separate) of the robot face and a context. On one hand, a simultaneous presentation might enhance the degree of the cue integration from either single-modality or multi-modality. On the other hand, a separate presentation might reduce the degree of the cue integration.

Because the above consideration of the cue integration explanation suggests that it may be better supported by the data than the emotional embodiment explanation, or the mood effect explanation, or the 'robot interpretation' explanation, more time will be devoted to its discussion.

From a cue integration perspective, it remains unclear how, when confronted with congruent or conflicting emotional cues, subjects combined them to reinterpret the emotional content of the robot expressions. Did they evaluate both sources separately or did they integrate them to make a final decision? Aronson's (1997) cognitive dissonance theory may be relevant to the question. According to cognitive dissonance theory, cognitive dissonance is the feeling of tension that comes as result of holding two conflicting thoughts at the same time. This theory argues that dissonance is a negative drive state that encourages attitude change in order to restore consistency, which results in attempts to devise ways of changing one or both cognitions that are in conflict to reduce the inconsistency and return to consistency.

Cognitive dissonance theory is supported by two computer studies that are related to the present experiments. In one study, Nass et al. (2001) investigated the mismatching of emotional voices and content, and found out that the consistency between content and emotional voices is important as participants reported liking content more when voice emotion and content emotion were matched. In the other, Creed and Beale (2008) investigated how mismatched facial and audio expressions were perceived when subjects viewed an animated avatar face accompanied by a female human voice. They found that when neither the visual nor audio dimension was dominant subjects did not attempt to make the expression consistent along one of the dimensions, but both. Therefore subjects attempted to make mismatched expressions consistent on both the visual and audio dimensions of animations with the result that their perception of the emotional expressions became confused. For example, subjects rated animations consistent with a happy or warm face, or with a happy or warm voice as more engaging, warm, concerned and happy. However, these two studies involved situations in which neither channel was dominant, and subjects made efforts to restore consistency by changing both of the conflicting dimensions. Aronson's (1997) cognitive dissonance theory also suggests that in cases where one of the channels is dominant over the other, subjects only make the cognition consistent along the dominant dimension. This prediction is supported by the studies of avatar faces (Noël et al. 2009; Mower et al. 2008; 2009) where the audio channel overrode the visual channel, and subjects attended to the avatar voices and based their interpretation on them.

Cognitive dissonance theory and the above studies suggest that how subjects try to restore consistency depends on which channel is more dominant. Channel dominance also influences how subjects combine congruent or conflicting emotional cues to reinterpret either channel. In human face studies, Niedenthal et al. (2006b) suggested that when facing discrepant cue combinations, observers will either evaluate both sources separately and choose the emotional meaning of the source that is the clearest and easier to access, or integrate both sources before the final decision, which make both sources susceptible to reinterpretation. Fernandez-Dols and Carroll (1997) defined vulnerability to reinterpretation as the extent to which the interpretation of one source influences the interpretation of the other. As enumerated by Niedenthal et al. (2006b), there are two types of interpretational efforts observers make to integrate the conflicting cues into one plausible emotional judgment. In the first type, a reinterpretation of the context requires observers to add new situational information in order to make the context consistent with the facial expression. This type of reinterpretation is most likely when the situational information is vague and general in tone and when it lacks concrete, specific information (e.g., to describe an event, a context is presented through a short stereotypical one sentence). In the second type, the reinterpretation of the facial cues requires observers to integrate conflicting situational cues to interpret the facial expression as being controlled by situational and social demands, such as not crying in public, rather than as reflecting the person's true emotion. This is more likely to occur if the facial expression is ambiguous. When neither source is dominant, both sources are susceptible to reinterpretation which makes the reinterpretation bidirectional; in contrast, when one of the sources is dominant, the interpretation of the dominant source influences the interpretation of the other which makes the reinterpretation one directional.

Based on the above explanations, the prediction can be made that to make one plausible emotional judgment about either the robot face or the context, subjects either evaluate the two sources separately or integrate the cues from the two sources together. And integration of the cues leads to the interpretation of one source being influenced by the interpretation of the other. The vulnerability of the robotic facial cues or the emotional cues of the context to reinterpretation depends on how dominant each source is. In other words, if the robot face is less dominant than the context, then the reinterpretation of the robot face is likely to be susceptible to the reinterpretation of the context; if the context was ambiguous, and the robot face was dominant, the context would be reinterpreted to be consistent with the face; if neither one is dominant, the reinterpretation may be bidirectional: the reinterpretation of the context is vulnerable to the reinterpretation of the robot face, and vice versa.

Table 4.28 Recognition accuracy (%correct) of each emotional context in the 11 experiments

Context	Simultaneously presented with the robot face on a first viewing		Simultaneously presented with the robot face on a second viewing		Presented preceding the robot face		Mean and standard deviation of each context in all the experiments	Mean and standard deviation of the robot face presented with the context
	Congruent with the robot face	Incongruent with the robot face	Congruent with the robot face	Incongruent with the robot face	Congruent with the robot face	Incongruent with the robot face		
Positive/Neutral BBC News	100	100	100	76.9	No data collected		M=94.23 SD=11.55	M=72.3 SD=31.06
Negative BBC News	94.4	83.3	84.6	100	No data collected		M=90.58 SD=8.00	S=61.63 SD=29.18
Positive Affective Pictures	100	96	100	96	87.5	71.4	M=91.82 SD=10.99	M=77.82 SD=16.92
Negative Affective Pictures	96	100	100	100	100	100	M=99.33 SD=1.63	M=71.48 SD=34.59
Positive Classical Music	100	73.3	86.7	100	92.9	85.7	M=89.77 SD=10.16	M=80.95 SD=25.41
Negative Classical Music	93.3	53.3	93.3	66.7	50	50	M=67.77 SD=20.72	M=86.53 SD=13.90
Positive Film Clip	93.3	93.3	85.7	87.5	76.9	84.6	M=86.88 SD=6.15	M=71.58 SD=14.92
Negative Film Clip	100	100	100	100	100	100	M=100 SD=0	M=73.05 SD=18.67

This explanation can be considered further with reference to Table 4.28. This table displays the standard deviation of subjects' recognition of each context or the robot face in all of the different manners of presentation (i.e., simultaneous presentation on a first viewing or second viewing, and separate presentation). It can be argued that these standard deviations reflect the vulnerability of the robot face or the context to reinterpretation: the bigger the standard deviation, the stronger vulnerability of the robot face or of the context to reinterpretation. Likewise, the smaller the standard deviation is, the more dominant the robot face or the context can be assumed to be. Take the negative affective pictures for example their recognition accuracies remain more or less the same (Mean=99.33%) no matter which manner of presentation is used for them, which results in their standard deviation (SD=1.63) being the smallest among all the contexts. This small standard deviation indicates that among all the contexts, the negative affective pictures are the least susceptible to reinterpretation in terms of the robot face, which makes the negative affective pictures the most dominant context. At the same time, the standard deviation (SD=34.59) of the robot face presented with the negative affective pictures has the largest standard deviation in the table which indicates that the robot face was most vulnerable to the reinterpretation in the light of the negative affective pictures. In contrast, the positive affective pictures are less dominant than the negative affective pictures as a context (their SD=10.99 while the robot face's SD=16.92). Even so, a contextual effect was found for the affective pictures in the manner of simultaneous presentation with the robot face (on a first viewing and a second viewing), and they are more dominant than the robot face in attributing the robot expressions to its emotions.

Compared with the affective pictures, the News recordings are less dominant (SD=11.55 for positive News and SD=8.00 for negative News) but the reinterpretation of the robot face is very susceptible to the reinterpretation of the News recordings (SD=31.06 for robot face presented with positive News and SD=29.18 for robot face presented with negative News). This is consistent with the finding that the News recordings have a situational influence on the robot expressions and fits with the conclusion that dominant over the robot face in attributing the robot expressions to its emotions. Moreover, a contextual effect was found for the musical pieces but they are not dominant over the robot face. In fact, such effect is bidirectional: the musical pieces affected the reinterpretation of the robot face and reversely the robot face influenced the reinterpretation of the musical pieces. But it is difficult to judge from Table 4.28, that the SD of musical pieces or the robot face presented with them can reflect that both the musical pieces and the robot face are susceptible to reinterpretation. Differently, not only the film clips seem not to be vulnerable to reinterpretation (SD=6.15 for positive film clip and SD=0 for negative film clip) by the robot face, but also the robot face seemed not be vulnerable to reinterpretation (SD=14.92 for robot face presented with positive film clip and SD=18.67 for robot face presented with negative film clip) influenced by the film clips. This is different from other cases that at least one of the two sources is susceptible to reinterpretation. Neither of the two sources was susceptible to reinterpretation when the robot face was presented with the film clips, probably resulting in that subjects have evaluated the two sources separately. This perhaps could explain why no contextual effect was found for the film clips.

In other words, when a robot face is reinterpreted and influenced by the interpretation of a context, a contextual effect of the context on the robot face will possibly occur. And when both the robot face and the context are susceptible to reinterpretation, such contextual influence is likely bidirectional. And it can be assumed that the explanation of people’s interpretation of the robot’s expressions as though there was a response to the context is related to the cue integration explanation to some extent. For instance, the robot’s responding to the context explanation could enhance the bidirectional reinterpretation – the vulnerability of the context to its reinterpretation is more influenced by the robot face, and the vulnerability of the robot’s face to a reinterpretation that is influence by the context is enhanced.

However, explaining the reasons for the absence of a contextual effect of the film clips as a surrounding context (i.e., in experiments Video1 and Video2) seems to be more challenging. Three possible explanations for the absence of a contextual effect can be proposed. The first one is that it is a cue integration issue, but it was too difficult for the subjects to combine emotional cues from multi-modality such as film. The second one is that the film clips needed too much attention with the result that the subjects did not pay enough attention to the robot head. The third one that, the choice of film clips used in these experiments was not good (especially the positive one). The second of these explanations is argued to be less likely, since subjects were told by the experimenter that attention should be paid to the robot head at the same time as they watched the film. If the third explanation is correct, then it should follow that, a context effect could be obtained if a different positive film clip was used that was recognized by more subjects as being positive. In experiments Video1, Video2 and Video3, subjects' recognition accuracy of the positive film clip (averagely less than 90%) was lower than that of negative film clip (averagely 100%), which indicated that the positive film clip was of a poor choice, as opposed to the negative film clip. However, no hard evidence was collected that clearly indicates that the positive film clip was a poor choice. Therefore, the first explanation remains the most likely.

If the absence of a context effect is due to a cue integration issue as in the first explanation above, this might help to explain the gender effect observed in Video1, as illustrated in Table 4.27. The prominent content of the positive film clips is about a woman simulating an orgasm in front of a man in a restaurant. Female subjects might have felt more uneasy or embarrassed about this, as compared to male subjects. Therefore it was more difficult for the female subjects to combine emotional cues from multi-modality (video to robot face). As a result they showed a greater tendency to evaluate the emotional valence of the video and the robotic expressions separately and their emotional judgments were less influenced by the video, compared to the male subjects. This account is supported by the observation that, female subjects (23 out of 27) recognized the sequences of the robotic expressions more accurately than the male subjects (20 out of 33). However, such a gender effect was only observed in Video1, but not in Video2 or Video3, which in turn weakens the explanation.

Despite some the uncertainties discussed above, this chapter establishes that three particular forms of context, namely, speech audio, still images, and music do have the ability to exert a contextual effect. This strongly implies that the contextual effects found here with a robot head are also similar to the contextual effects found with human faces and avatar faces. This also leads to the conclusion that the robot face was viewed in the same way as human faces, since human emotions are both biologically based and socially constructed. Moreover, the contextual effects found here are consistent with the argument in Social Constructionist Theories (see Russell et al. 2003) that attributing a specific emotion to the sender becomes more complex when it occurs in an emotional situation, since even when the stimuli are facial expressions of basic emotions, the context which they occur (Carroll and Russell, 1996) can color the attribution. It is therefore argued that Social Constructionist Theories also apply to the recognition of the emotional expressions of a robot. The theory of embodying emotion may also help us to understand the emotional congruence observed in this chapter in the recognition of the robotic expressions.

However, there are still a few unanswered questions, such as the following:

1. No contextual effect was found for film clips. Why did film clips seem to be an exception to all the surrounding contexts? Was this because the film clips, especially the amused/happy film clips were badly chosen?

2. As was described at the outset of this chapter, the FACS was applied to program the synthetic robot facial expressions, and all the emotional materials were chosen either based their content (e.g. the recorded BBC News) or from some lists provided in some previously published studies (e.g. the affective pictures, the classical music and the film clips). However, it would be good to obtain further scientific evidence to support the correct categorisation of the answers assumed by the experimenter to be right.

Such questions will be answered as much as possible and further evidence sought by means of a further experiment and further analysis of the experimental methodology of the 11 experiments in next chapter.

Chapter 5: An Analysis of the Methodology

11 experiments have been described in the previous chapter, with a few questions remaining, such as, were the five different kinds of materials (i.e., the robotic facial expressions, the recorded BBC News, the affective pictures, the classical music, and the film clips) chosen with proper valence? And, how to explain the reasons for the absence of a contextual effect of the particular pair of film clips? A further experiment subsequent to the previous 11 experiments will be described firstly in this chapter to answer these questions. This chapter will provide an analysis of the methodology used in the previous 11 experiments, by taking into account the findings of the further experiment, and identifying the factors one by one that may have contributed to the contextual effects found in the previous chapter.

5.1 A Further Experiment

5.1.1 The Motivation of the Experiment

This experiment was conducted after the previous 11 experiments, with the aim to seek further confirmation about the recognition accuracy and the emotional valence of the materials when viewed separately, and of the robot head sequences when viewed without context. In other words, by comparing the recognition rate and the emotional valence of the robot expression sequence and each context when viewed separately, not only can further scientific evidence for the right answers provided by the experimenter be obtained, but also the reasons for finding or not finding a contextual effect may be found.

5.1.2 The Experimental Conditions

The two sequences of facial expressions (positive and negative) of the emotional robot used in this experiment were the same as those used in the first experiment. The film clips, the affective pictures, the recorded BBC News, were as the same as those used in the earlier 11 experiments. The classical music, was as the same as that used in the fourth experiment which lasted for about 3 minutes. In this experiment, each subject first filled in a questionnaire about how they perceived the robot's simulated emotions and each emotional material in isolation (the robot facial expressions will be presented firstly, followed by, in a random order one after another, the recorded BBC News, the affective pictures, the classical music, and the film clips). The emotional materials viewed by each subject were either all positive or all negative. They rated the emotional valence of each emotional material (i.e., they saw all the materials, and then rated them all) by using 9 point Self-Assessment-Manikin scale adapted from (Lang et al. 2008).

Therefore, subjects were divided into two groups. Subjects in Group 1 (or Condition 1) watched or listened to five different positive emotional materials, while subjects in Group 2 (or Condition 2) watched or listened to five different negative emotional materials (a sequence of robot facial expressions was included in either condition).

5.1.3 Experimental Procedures

Subjects in this experiment read a similar information sheet as instructions and filled in a similar questionnaire as provided in the experiment Speech1 (see Appendix Ⅱ: The Questionnaire of the Experiment Speech1). For instance, the overview of the project procedures provided in the information sheet of this experiment was presented as follows:

Firstly, to warm you up, you will be shown a few different facial expressions of an emotional robot. Secondly, you will see a sequence of facial expressions of the emotional robot, and then to fill in a short questionnaire. In the third part of the study, you will be asked to fill in a short questionnaire after viewing the other four emotional materials (the record BBC News, the affective pictures, the emotional classical music and the film clips) one after another in a random order. Finally, you will be asked to fill in a final short questionnaire regarding the comparison of the emotional valence (from positive to negative) of these emotional materials. During this project, please keep in mind that these emotional materials including the robot emotions are not corresponding to each other and you are supposed to treat each of them as an isolated individual. There will be no time for an explanation of the project at the end, further details or discussion will have to take place via emails.

Then, he/she went through the following procedures:

Warm up: Six different facial expressions of the emotional robot CIM were shown to all subjects.

Experiment Part One: Subjects watched a 3-minute sequence of facial expressions (either positive or negative) of the emotional robot.

Responses Part One: After the robot emotions, subjects were asked to answer the following questions:

1. As a total impression, please select what kind of emotion (affect) you think the robot was feeling from the following given choices.___________

A: Positive Affect B: Neutral Affect C: Negative Affect

Experiment Part Two: Subjects watched or listened to the other four emotional materials, namely the BBC News, the Film Clips, the Classical Music, and the Affective Pictures one after another in a random order. All the emotional materials were either all positive or all negative, depending on the facial expressions of the robot they had seen in the first part was positive or negative (e.g. if the robot face showed positive facial expressions, then all the other four emotional materials would be positive).

Responses Part Two: Every time subjects finished watching or listening to an emotional material, they were asked to answer the following questions:

As a total impression, please select what kind of emotional material you think you were viewing from the following given choices. ___________

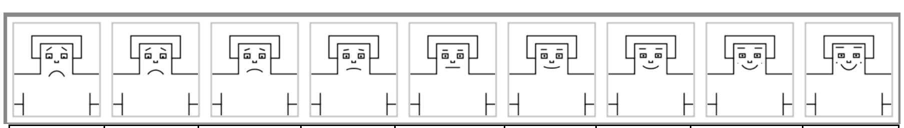
A: Positive Material B: Neutral Material C: Negative Material

After part two, subjects were asked to rate the emotional valence of each emotional material by the following questionnaire:

Please use the following Self-Assessment-Manikin scales to rate the emotional valence of each of the emotional materials. Circle any number underneath any of the figures (1 indicates most negative and 9 indicates most positive):

And an example of the five questions (subjects needed to circle their answers for the Synthetic Robot Emotions, the Recorded BBC News, the Affective Pictures, the Classical Music, and the Film Clips, respectively) is shown as follows:

Please circle your answer for the Synthetic Robot Emotions:



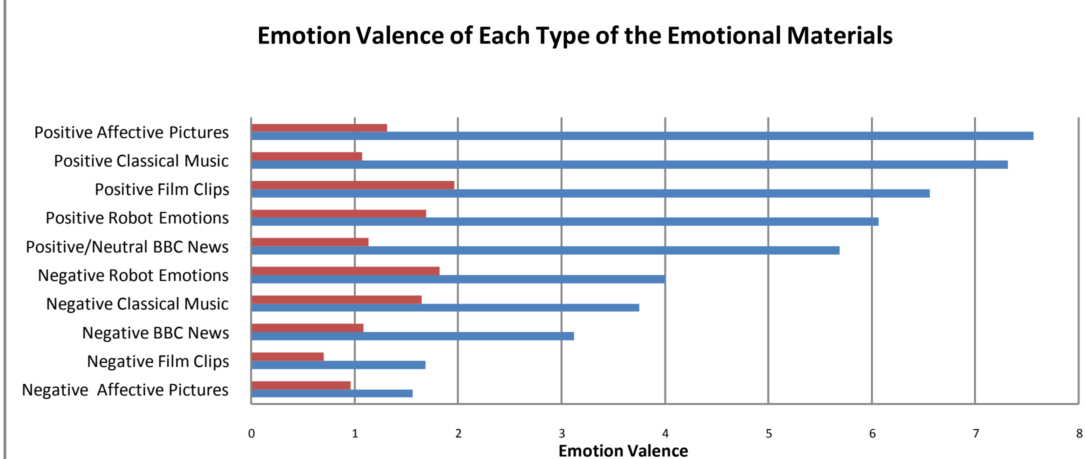
1	2	3	4	5	6	7	8	9

5.1.4 Experimental Results

32 volunteers with average age of 26.3 participated in this experiment. These 14 male participants and 18 female participants, had various nationalities, as

there were 17 British, 11 Chinese, 2 other European participants (1 Swedish, and 1 Polish), 1 Indian, and 1 Caribbean. Thus, there were 16 subjects in each group. They were partially recruited from the university mailing list as described in the first experiment, and partially recruited from the leaflets as described in the fourth experiment. Care was taken to ensure that participants in this experiment had not taken part in the earlier experiments. And each of them was paid £5 for each section. There were 16 subjects in each group, or each condition.

The responses to the questionnaire administered after the rating task were analysed (see Fig. 5.1).



	Negative Affective Pictures	Negative Film Clips	Negative BBC News	Negative Classical Music	Negative Robot Emotions	Positive/ Neutral BBC News	Positive Robot Emotions	Positive Film Clips	Positive Classical Music	Positive Affective Pictures
Standard Deviation	0.96393	0.70415	1.08781	1.65328	1.82574	1.13835	1.69189	1.96532	1.07819	1.31498
Mean Valence	1.5625	1.6875	3.125	3.75	4	5.6875	6.0625	6.5625	7.3125	7.5625

Fig. 5.1: Emotion Valence of Each Type of the Emotional Materials including the robot face rated by all subjects

Serving as a further experiment, this experiment provided some important indications for the previously conducted 11 experiments.

First of all, from Fig.5.1, it is evident that the contextual materials used in the preceding experiments were of the appropriate valence. As when viewed without any surrounding emotional context, the robot's expressions were recognized as negative in the negative sequence (average rating of 4), and positive in the positive sequence (average rating of 6.25). This reinforces the earlier claims about the robot's expressions, which were developed based on the FACS.

Moreover, the four emotional materials, which were either chosen from the lists of what some previous researchers have recommended (the affective pictures, the film clips, and the classical music), or selected from the easily obtained BBC News, were rated as expected – appropriately recognized as positive or negative. The data of the mean valence rated here supports the use of emotional contexts that are either positive or negative (according to the 9-point Self-Assessment-Manikin scale used in the experiment, 1 indicates most negative and 9 indicates most positive, while 5 indicates most neutral). It validates the selection of affective pictures, music and film clips based on previous lists, and also shows that the BBC news was viewed as intended (positive in the positive/neutral condition, and negative in the negative condition). Interestingly, among all the five emotional materials, the affective pictures were viewed as the most positive and the most negative, whilst classical music could be very positive but not very negative, film clips could be very negative but not very positive, the recorded BBC news could be medium negative but just a bit positive, and the robot emotions were neither very positive nor very negative.

Secondly, although no statistical evidence was collected, the data shown in Fig 5.1 could imply that the News recordings were unlikely to induce strong moods in the subjects because of their weaker emotional valence. In contrast, the affective pictures may have had the ability to color subjects' mood states, judging by the strength of the affective valence (or the pleasure) of each material obtained by the 9-point Self-Assessment-Manikin scale. This deduction is made also based on the evidence that both of the 3mins long classical music and the film clips were observed to have strong influence on subjects' mood states while no mood effect was observed for the robot expressions, in the previous 11 experiments as described in Chapter 4. In the previous 11 experiments, a contextual effect was observed for both the News recordings and the affective pictures, and it was proposed that mood effect was one of the possible explanations for their contextual effect, although no statistical evidence was collected to support such mood explanation. Therefore Fig. 5.1 could be considered to provide some evidence that supports the suggestion that a mood effect was one of the factors involved in the contextual effect of the affective pictures, whilst not supporting the same suggestion for the News recordings.

Thirdly, as can be seen in Table 5.1, subjects were able to recognize robot emotions with considerable high accuracy of 75% in both conditions. It can be seen in Table 5.2 that subjects were perfectly good at identifying the emotional content of Positive Classical Music, Positive/Neutral BBC News, Negative Film Clips, and Negative Affective Pictures with a perfect accuracy of 100%; almost perfectly good at identifying the emotional content of Positive Affective Pictures and Negative BBC News with a very high accuracy of 93.75%; relatively good at identifying the emotional content of Negative Classical Music considerable high accuracy of 75%; but poor at identifying the emotional content of Positive Film Clips with a low accuracy of 56.25%. It is worth mentioning that among these who incorrectly perceived the Positive Film Clips, 5 out 7 subjects were female (4 of them were Chinese).

Table 5.1 Percentage of three possible labels (items in the first row) chosen to match the displayed sequence of facial expressions (positive or negative).

%match	Positive Affect	Neutral Affect	Negative Affect	%correct
Positive Robot Affect (Condition 1)	75	0	25	75
Negative Robot Affect (Condition 2)	12.5	12.5	75	75

Table 5.2 Percentage of three possible labels (items in the first row) chosen to match four other emotional materials.

%match		Positive Material	Neutral Material	Negative Material	%correct
Condition 1 (Positive Emotional Materials)	Positive Affective Pictures	93.75	6.25	0	93.75
	Positive Film Clips	56.25	25	18.75	56.25
	Positive Classical Music	100	0	0	100
	Positive/Neutral BBC News	100	0	0	100
Condition 2 (Negative Emotional Materials)	Negative BBC News	0	6.25	93.75	93.75
	Negative Classical Music	12.5	12.5	75	75
	Negative Film Clips	0	0	100	100
	Negative Affective Pictures	0	0	100	100

Fourthly, these results help to distinguish between some possible explanations of the reasons for an absence of contextual effects from the film contexts. Three possible explanations for the absence of a contextual effect for the film clips were considered in the last section of Chapter 4. It was deemed unlikely that subjects did not pay enough attention to the robot head. And it was argued that the cue integration explanation remained the most likely as no hard evidence was collected to clearly indicate that the positive film clip was a poor choice. However, in this experiment, a Fisher's exact test in PASW 18 indicated the accuracy of subjects' perception of the positive film clip (a relatively lower accuracy, 9/16 (56.25% correct, 43.75% incorrect)) was significantly different from that of the negative film clip (a perfect accuracy, 16/16 (100% correct, 0% incorrect)) with two-tailed p=0. 0068. This provides evidence that the positive film clip used in the previous experiments was indeed a poor choice, compared to the negative film clip. This lends more support for the conclusion that no contextual effects were found for the film clips was also due to that positive film clip were not a good choice.

And it was proposed that, unlike the other emotional materials, the particular pair of film clips did not have a significant contextual effect, because when witnessing the positive synthetic robot emotions and positive film clip simultaneously, subjects did not see the congruence between the two positive emotional materials. In other words, they did not attempt to integrate the emotional cues from the two emotional materials, but evaluated both materials separately. This explanation is supported by the observation that the recognition accuracy of the synthetic robot facial expressions accompanying the positive film clip in either condition in the previous 11 experiments was more or less as the same as in this experiment in which subjects watched the positive synthetic robot emotions without any surrounding context or any preceding context. In contrast, the negative film clip seemed to make a difference between subjects' perception of congruent and conflicting robot facial expressions. For instance, in Video1, subjects correctly perceived congruent robot facial expressions by 93.3%, as opposed to only 40% when confronted with incongruent ones. As Niedenthal et al. (2006b) pointed out, when confronted with combined emotional cues from two sources, observers will either evaluate them separately without any integration or integrate the emotional cues from both sources to evaluate the emotional meaning of each of the sources before the final decision. It is possible that in the case of film clips, subjects found it too difficult to integrate the emotional cues from the positive film clips and the positive synthetic robot emotions so that they had to evaluate them separately.

Fifthly, there was no gender effect on or cross-cultural difference in the accuracy of subjects' perception of the film clips (positive film clip with negative film clip together). As a Fisher's exact test in PASW 18 indicated the accuracy of subjects' perception of the film clips was not significantly different between male subjects (a relatively higher accuracy, 12/14 (85.7% correct, 14.3% incorrect)), and female subjects (a high accuracy, 13/18 (72.2% correct, 27.8% incorrect)) with two-tailed p=0. 426. And a Fisher's exact test in PASW 18 indicated the accuracy of subjects' perception of the film clips was not significantly different between western subjects (a relatively higher accuracy, 16/19 (84.2% correct, 15.8% incorrect)), and non-western subjects (a high accuracy, 9/13 (69.2% correct, 30.8% incorrect)) with two-tailed p=0. 401.

5.1.5 Conclusion and Discussion

This section describes a further experiment conducted after the 11 experiments as described in Chapter 4. It examines how subjects perceive the robot's simulated emotions when presented without an accompanying context. It also examines how each of the emotional contexts is rated when viewed without the presentation of the sequence of the robotic facial expressions (although subjects were told to treat each of the emotional contexts as being isolated from the robot emotions, the robot head remains static in front of the participant). As such, it provides scientific validation that the robot's expressions and the contextual materials used in the preceding experiments were of the intended valence.

The experiment reinforces the claims about the robot's expressions, which were developed based on the FACS, and supports the selection of the recorded BBC News, affective pictures, musical pieces and film clips. Secondly, it confirms that when viewed individually, each sequence of robot facial expressions or almost all the contextual materials were correctly recognized with accuracy of at least 75% as intended, although subjects were poor at identifying the emotional content of Positive Film Clips. The accuracy of at least 75% actually serves as a baseline since subjects normally did not achieve a recognition rate of robot facial expressions as high as 75%, when confronted with conflicting cues from the context and the robot face. Thirdly, by comparing the recognition rate and the emotional valence of each context, a conclusion can be drawn that the particular pair of film clips seems to be an exception to all the surrounding contexts. It was suggested that possibly this is because positive film clip was indeed a poor choice, and subjects found it too difficult to integrate the emotional cues from the positive film clip and the positive synthetic robot emotions so that they had to evaluate them separately.

However, in order to provide a more thorough answer to the question about which explanation for the absence of a contextual effect for the film clips is best, the next section identifies and discusses possible factors contributing to the contextual effects found in the previous 11 experiments.

5.2 Possible Factors Contributing to the Contextual Effects Found in the Previous 11 Experiments

In the 11 experiments described in Chapter 4, it seemed that the particular pair of film clips was an exception to the surrounding contexts. Compared with other surrounding contexts (e.g. the recorded BBC News, the affective pictures, and the classical music), when a film clip was presented with the synthetic robot facial expressions to the subjects at the same time, whether the film clip was congruent with the emotional valence of the robot emotions or not, did not affect subjects' perception of the synthetic robot facial expressions significantly. This suggests that different kinds of context may exert different kinds of effect on the recognition of the robot's expressions. To analyze what influences the contextual effects on subjects' recognition of the robotic facial expressions, some possible factors contributing to the contextual effects found in the previous 11 experiments will be identified. The following analysis will focus on the choices of the contexts, the rating scheme, the mood effect, and the manner of presenting the robot face and the context. With respect to the choices of the contexts, their content, their source clarity, and their perception difference between single modality and multiple modalities, will be considered. As for the manner of presentation, the style of the presentation of a robot face and a context, and the contrast effect were taken into account.

5.2.1 The Choices of the Contexts

It was concluded earlier in this chapter that the positive film clip was indeed a poor choice - compared with the positive film clip, the negative film clip not only were recognized better, but also had the ability to make a difference between subjects' perception of congruent and conflicting robot facial expressions. And it was therefore argued that different kinds of context may have different abilities to exert a contextual effect on the recognition of robotic expressions. Consequently, care should be taken when selecting contexts to conduct experiments in which a contextual effect can be found. The following three aspects show what to focus on when selecting contexts:

A. The Content of the Contexts

As mentioned before, amongst all the five different kinds of emotional material, the positive film clip was the most ambiguous material as perceived by the subjects. Fogg's (2003) Prominence-Interpretation Theory can be used here to explain why subjects judged the positive film clips as relatively ambiguous, as opposed to other emotional materials. Fogg's theory evolved from work on Web credibility, but some of the principles of the theory also apply in my experiments involving film clips. The theory suggests that there are two primary components that influence the perceptions of website credibility – (1) Prominence – an element's likelihood of being noticed, and (2) Interpretation – a person's judgment about an element under examination.

Fogg suggests that there are numerous factors that influence both Prominence and Interpretation. For instance, Fogg highlighted five primary factors that affect Prominence – (1) Involvement of the user (motivation to process & ability to process, e.g., if users have a high level of motivation to find out answers on a Web site, they will notice more things about the Web site); (2) Content of the Web site (news, health, sports, etc.); (3) Task of the user (seeking information, seeking amusement, making a transaction, etc.); (4) Experience of the user (with Web, with subject matter, etc.); (5) Individual differences (need for cognition, learning style, literacy level, etc.). Fogg suggests that the most important factor should be a user's involvement.

With regard to Interpretation and its influence on credibility, Fogg lists the most important factors he viewed as follows– (1) Assumptions in user's mind (culture, past experiences, heuristics, etc.); (2) Skill/knowledge of user (level of competency in subject matter, etc.); (3) Context (e.g., environment, norms, expectations); (4) User goals (e.g., to find out information, to make a transaction, etc.). Fogg suggests that both culture (e.g., some people interpret the same Bible verse in a prominent place of a Web site more positively than others) and context (e.g., people will find a popup more distracting and interpret it more negatively if they are more in a hurry to look for the best airfares online) play roles in Interpretation.

Presumably, some of the principles of Fogg's theory also apply in my experiments. From a prominence perspective, some of the subjects might not find any elements of the positive film clip amusing, or something to laugh at. It is possible as they might lack of sense of humor or may lack the necessary linguistic skills (e.g., if their English level is not good enough to catch the last sentence spoken by the woman in the amused (happy) film clip). From an interpretation perspective, due to the cultural background of some subjects, especially some Asian female subjects, even if they have noticed some elements that are supposed to be amusing, they might feel uneasy or embarrassed (i.e., the prominent content of the positive film clips is about a woman simulating an orgasm in front of a man in a restaurant). As a result of these two perspectives, some of the subjects might judge the amused (happy) film clip as a neutral or even a sad film clip.

B. The Source Clarity of the Contexts

According to Niedenthal et al. (2006b), source clarity can be defined as the amount of emotional information provided by each of the two types of information (e.g., human facial expressions and human emotional voices) and can be comprised of three components, namely source ambiguity (the degree of a source corresponding to one specific emotion rated by the judges), source complexity (the number of different emotions attributed by different judges to the same source), and source intensity (the perceived intensity of the emotional message). They conclude that, the clearest source can determine the judgments of the meaning of an emotional expression in its context. As suggested by them, source clarity can be manipulated or controlled in the confines of laboratory, for instances, by increasing or decreasing the ambiguity or intensity of human facial expressions, or by heightening or weakening the salience of context information, which may bias the relative weight of facial and situational cues in emotional judgments.

In the studies presented here, the two types of information are robotic facial expressions and surrounding emotional contexts. With regards to source ambiguity, it has been shown in the experiment as described in this chapter that, when viewed individually, each sequence of robot facial expressions or each contextual material was correctly recognized with an accuracy of at least 75% as intended, although subjects were poor at identifying the emotional content of Positive Film Clips with an accuracy of 56.25%. In other words, amongst all the five emotional materials, positive film clips are the most ambiguous material. If Niedenthal et al.'s conclusion is correct, the reason that a contextual effect was not observed for the Positive Film Clips might be that - subjects found it too difficult to choose the Positive Film Clips as the clearest source to determine the emotional judgments of the robotic facial expressions accompanied by the Positive Film Clips. Also in the experiment as described in this chapter, the data of the rated mean valences supported the use of the sequence of robotic facial expressions or emotional contexts that are either positive or negative. In more detail, amongst all the five emotional materials, the affective pictures could be either most positive or most negative, classical music could be very positive but not very negative, film clips could be very negative but not very positive, the recorded BBC news could be medium negative but just a bit positive, and the robot emotions were neither very positive nor very negative. Consequently, the source intensity varies among all the five emotional materials.

As for the source complexity, the five emotional materials vary in terms of the content of emotions. Subjects could infer emotions from the robotic facial expressions, as they were based on the FACS. For the recorded BBC News, the subjects could not infer any emotion from the audio as the News was broadcasted in neutral tone, but could infer emotional information from the words describing the events. For the classical music, the subjects could not infer any emotion from the audio either, but their mood states could be colored by the classical music. For the affective pictures, subjects not only could infer emotions from pictures of human beings, but also could infer emotions from pictures of animals as many other species also have emotions and facial expressions of emotions are universal (Ekman, 2009). Of course, some strong emotional content could be also inferred from the affective pictures that of tragedies etc. For the film clips, subjects could infer emotions not only from human faces in the visual channel, but also from emotional speech in the audio channel. Again, strong emotional content could be also inferred from the film clips as they told the subjects amused or sad stories. In terms of the source complexity, it seems that the film clips are the most complex one among the four contexts, followed by the affective pictures or the BBC News. As a result, the classical music remains the least complex one.

After analyzing the source clarity in three aspects respectively, it can be arrived at that among the four contexts, the affective pictures are the clearest one, followed by the BBC News, and the film clips are the least clear one, followed by the classical music. Even though this is not reasoned on a statistical basis, it might explain why the relative weight of the surrounding context and the robot face on the emotional judgments of the robot face were different. This inspires a way of examining the changes of the relative weight in the future, by modifying the source clarity of the surrounding contexts having effects on the robot face (i.e., the recorded BBC News, selected affective pictures, and the musical pieces).

C. The Difference Between One Modality and Multiple Modalities

The subjects in the 11 experiments perceived the synthetic robot expressions and the contexts through the same visual channel or through multi-modality, depending on what the context was. For example, when the surrounding context was the affective pictures, the perception was in the same visual channel; when the surrounding context was the recorded BBC News or the classical music, the perception was in the multi-modality that of visual-audio; and when the surrounding context was the film clips, differently, the perception was in the multi-modality that of visual-audio combined with video.

As speculated in Chapter 4, perceiving emotional cues through the same visual channel should be easier than perceiving them through multi-modality. However, such speculation was not on based on empirical data. This is because contextual effects were observed not only when subjects perceived the emotional cues from the same channels (e.g., visual channel (still images) with visual channel (robot face) in Image1 and Image2), but also when they perceived them from two channels (e.g., visual-audio in Speech1, Speech2, Music1 and Music2). The fact that no contextual effect was found for the film clips when the perception was in the multi-modality, might support such speculation.

Nevertheless, care should be taken when conducting experiments involving perceiving emotional cues through multi-modality. It should be made sure that participants in such experiments pay enough attention to the robot head as well as the context.

5.2.2 The Rating Scheme

A forced choice rating scheme (e.g., question 1 and 2 described in the response section) was used in all the experiments to rate not only the sequences of robotic expressions but also the emotional materials. Using a forced choice rating scheme to rate the sequences of robotic expressions was inspired by The Media Equation (Reeves and Nass, 1996). According to The Media Equation, good versus bad is a primary evaluation of mediated experience, thus the evaluation of Positive/Negative applies to media, and Bartneck et al. (2005) extended the "Media" to robots, since they often have an anthropomorphic embodiment and human-like behavior. In human face and avatar face studies (e.g, Noël et al. (2009)' study), a single facial expression was often paired with a voice to be presented to the subjects within a very short time period, e.g., a few seconds. In the studies presented in this thesis, each sequence of facial expressions was about three minutes long: almost the same time length as each type of the surrounding contexts. Therefore, by using a forced choice rating scheme, subjects' total impression after the long period presentation of the robot face and the context can be evaluated.

The studies presented in this thesis also benefited from using a forced choice rating scheme to rate the emotional materials. In the previous 11 experiments, subjects were only asked to distinguish positive materials from negative materials, which could probably result in a fair accessibility of emotion categories for both the robotic facial expressions and the emotional surrounding contexts. This is because, Niedenthal et al. (2006b) speculated that, as a result of people's insufficient experience of categorizing situations in terms of emotion categories, facial expressions are easier and faster to categorize in terms of discrete emotions than descriptions of situations. As an example, Fernandez-Dols et al.' (1991) experimental result supported their prediction that if participants were trained to categorize situations using simple emotion terms, then the weight of context in the judgment of emotional experience should increase.

At the same time, a forced choice rating scheme might push subjects' responses in a particular direction. It might be that contextual effects are found here precisely because a forced choice rating scheme was used. Future research could make use of a more flexible rating scheme to rate the sequences of robotic expressions (e.g., the Geneva Emotion Wheel), to see whether the findings in this paper also apply to some subtle facial expressions. At the same time, more emotion categories can be used to describe the emotional materials in the rating scheme for contexts, although this would be challenging.

5.2.3 The Mood Effect

As described in Chapter 4, it was evident that both the 3mins long classical music and the film clips were observed to have strong influence on subjects' mood states while the robot expressions were observed to have no such mood effect. And as mentioned before, although no statistical evidence was collected, it can be indicated that from Fig.5.1, the News recordings were unlikely to induce strong moods in the subjects. In contrast, the affective pictures may have had the ability to color subjects' mood states, judging by the strength of the emotional valence of each material. A conclusion can be drawn that different kinds of context may have different abilities to induce moods in subjects.

In some cases, the accompanying or preceding contexts could affect subjects' mood states, which suggested subjects' mood states could also play a role in coloring the contextual effects. However, as argued in the last section of Chapter 4, a mood effect cannot be considered to be a sufficient factor for a contextual effect. At the same time, it remains unclear whether or not a mood effect is necessary for obtaining a contextual effect. Future experiments could control subjects' mood states to see if a contextual effect can occur in the absence of a mood effect. If it did, it can be concluded that a mood effect is not necessary in obtaining a contextual effect.

5.2.4 The Manner of Presenting the Robot Face and the Context

After a context is selected, and the convincing facial expressions of a robot are created, care should be taken over how to present a robot face and a context (i.e., simultaneously or separately, on a first viewing or second viewing of the robot face, and whether subjects' have or have not seen the robot expressions before). Whether or not a contextual effect will be observed depends on the manner of presentation.

A. *The Style of the Presentation of a Robot Face and a Context*

Earlier studies with human faces (Carroll and Russell, 1996; De Gelder and Vroomen, 2000) and avatar faces (Creed and Beale, 2008; Noël et al. 2009; Mower et al. 2008; Mower et al. 2009) have shown that the way in which facial and contextual cues were combined (congruent or incongruent with each other), could affect observers' judgments about facial expressions of humans or avatars. Contextual effects were found in these studies. In order to examine whether such effects could also be obtained in studies with robot faces, the methodology used in these human and avatar studies was used in selecting the way in which the facial cues of a robot and the contextual cues were combined. And to see the circumstances under which such contextual effects are likely to occur, it was investigated whether the manner in which the context and the robot head are presented (simultaneously or separately) is important in obtaining such a contextual effect.

Evidence was observed that the manner in which the context and the robot head are presented (simultaneous or separately), and the way in which the facial cues of a robot and the contextual cues were combined (congruent or incongruent with each other), could affect subjects' judgments about synthetic robot facial expressions. Observers were better at recognizing the robot's expressions when they matched the emotional valence of accompanying surrounding contexts, than when they did not.

As was concluded in chapter four, a contextual effect only occurs in a simultaneous but not separate presentation of a robot face and a context and it neither depends on whether that context can color observers' own mood states nor depends on whether or not observers have seen an opposite valence of the sequence of robotic facial expressions before. Therefore, the style of the presentation of the robot face and a context is considered as the most important factor contributing to the contextual effects in this thesis.

B. *The Contrast Effect*

To see the circumstances under which such contextual effects are likely to occur, it was also investigated whether a second viewing of the robot face (observers have seen an opposite valence of robot facial expressions before the presentation of the robot face) can weaken a contextual effect. The idea of a second viewing of the robot face was generated after knowing that previously encountered facial stimuli can color the recognition of facial expressions of human beings.

As mentioned in Chapter 4, whether or not observers have seen an opposite valence of the sequence of robotic facial expressions before (i.e., a second viewing of the robot face), was not found to affect the occurrence of a contextual effect significantly. Although no experimental data was collected to support the contrast effect in the recognition of synthetic robot facial expressions, the contrast effect could be considered as a factor contributing to the contextual effects. According to Niedenthal et al. (2006b), the types of human facial expression that the observer has encountered previously may result in contrast effects (e.g., changing the attribution of the emotion category label). Therefore, collecting experimental data regarding the contrast effect could be undertaken in the future.

Note that subjects were asked to see all the six basic facial expressions of the robot in the warm up procedure of all the six studies, which might cause some kind of contrast effect in the later evaluation of Positive/Negative to the robot facial expressions. However, each of the six basic facial expressions was displayed for a very short time (less than 3 seconds), and according to the recognition rate in the warm up procedure of the first experiment, they are of a balance of positive (joy, surprise, and fear (most of subjects attributed fear to joy and surprise)) and negative (sad, anger, and disgust) facial expressions. Even if contrast effects were generated, the contextual effects found in all the six studies were still very strong.

5.3 Summary and Conclusion

This chapter has described a further experiment, in which further validating evidence was obtained that the robot's expressions and the contextual materials used in the preceding 11 experiments were of the intended valence. Generally the experiment provides scientific evidence that supports the correctness of the answers assumed by the experimenter to be right with the exception of the positive film clip. It seems that the positive film clip was indeed a bad choice, as opposed to other emotional materials when they were rated separately. These findings lead to the speculation that subjects found it too difficult to integrate the emotional cues from the positive film clips and the positive synthetic robot emotions so that they had to evaluate them separately, resulting in no situational influence of the positive film clip on the robot expressions (or vice versa). However, this does not necessarily indicate that video as a particular type of context does not have the ability of exerting a contextual effect as other types of context (i.e., speech audio, still image and music) do. It was observed in Video1 that whether or not the negative film clip was congruent with the emotional valence of the robot emotions, did affect subjects' perception of the robot expressions significantly.

To find out why it was not always the case that a contextual effect was found for each type of context in every specific situation, some possible factors contributing to the contextual effects found in the previous 11 experiments are identified. The choices of the contexts, the rating scheme, the mood effect, and the manner of presenting the robot face and the context, were considered. From a context perspective, it is concluded that different kinds of context vary in terms of their content, their source clarity, and their perception difference between single modality and multiple modalities. From a rating scheme perspective, it is argued that contextual effects are found here precisely because a forced choice rating scheme was used. The mood effect was argued not to be a sufficient factor for a contextual effect while more efforts should be made to find out whether it is a necessary factor for a contextual effect. And from a presentation perspective, the methodology used in previous human face and avatar face studies was used in selecting the way in which the facial cues of a robot and the contextual cues were combined. And such methodology was confirmed as effective since contextual effects on a particular robot face were obtained. A conclusion can be arrived at that a contextual effect only occurs in a simultaneous but not separate presentation of a robot face. Whether or not observers have seen an opposite valence of the sequence of robotic facial expressions before seems to make no difference, even though there might be perception difference between single modality and multiple modalities.

On the surface, it seems that video was an exception to the contexts as contextual effects were found for other three types of contexts. However, as mentioned above, video as a particular type of context still has the ability of exerting a contextual effect. Further experiments are needed to explore the possibility of using film clips to obtain a contextual effect. As the film experiments were not really valid, some new experiments will be proposed in next chapter to properly address the impact of film as a situational context for emotion recognition. A better choice of film clips will be used. Some factors influencing the occurrence of a contextual effect were identified in this chapter. The new experiments will take such complexities into account. For instance, a mood effect is identified as a factor, which could be controlled by exposing the participants to some stimuli to achieve a more consistent emotional state to begin with.

Chapter 6: Proposed New Experiments Addressing the Impact of Film Clips

As described in Chapter 4, four types of context were investigated in 11 experiments and three of them were observed to have situational influences on the facial expressions recognition of a particular robot head. And as validated in Chapter 5, participants in the experiment subsequent to the previous 11 experiments had no difficulties in recognizing these four types of context with the exception of the positive film clip which was not well recognised. In addition, evidence was obtained that demonstrated that all of the contexts had the intended valence. It was therefore argued in Chapter 5 that, by using a more controlled methodology than that used in the previous 11 experiments, some future experiments could be conducted to explore the possibility that a video context might exert a contextual effect under some circumstances. It was suggested that in order to obtain valid experimental results, a better choice of film clips would be made, and that factors influencing a contextual effect such as a mood effect should also be taken into account when conducting such experiment. This chapter will propose four new experiments (one preliminary experiment and three subsequent experiments) to address the impact of film clips.

The purpose of conducting a preliminary experiment is to select a piece of music that can be shown to induce neutral moods in subjects, and to select two pairs of film clips (one with the ability of inducing non-neutral moods to subjects, while the other with the ability of inducing neutral moods). This particular piece of music and the two new pairs of film clips will be used later in the three new experiments. The purpose of using neutral music in each new experiment is to neutralize subjects’ moods before showing them the film clip and sequence of robotic expressions, and to ensure that any mood effect that is obtained is the result of the film context and not their pre-existing mood. As for the purpose of using a pair of film clips that are likely to induce strong positive/negative moods in subjects in one new experiment and another pair of film clips that are only likely to induce neutral moods in another new experiment, it is proposed to investigate whether or not a mood effect is a necessary factor contributing to a contextual effect. A conclusion could be drawn that a mood effect is not a necessary factor contributing to a contextual effect if a contextual effect will be found for both of the two pairs of film clips in two different new experiments. And the complexities of the factors, such as the choices of the contexts, the mood effect, and the rating scheme, will be considered in the new experiments to enable the experimental methodology to be more controlled.

6.1. The Preliminary Experiment

6.1.1 The Motivation of the Preliminary Experiment

The main motivation of the preliminary experiment is to find some good choices of film clips. A further experiment described in Chapter 5 confirmed that the positive film clip used in the previous 11 experiments was indeed a poor choice, compared to the negative film clip. And it was argued that therefore subjects found it too difficult to integrate the emotional cues from the positive film clip and the synthetic robot expressions so that they had to evaluate them separately. This might explain why a gender difference on the perception of the robotic expressions was observed in the case of a simultaneous presentation of a film clip and a robot face (Video1), although no contextual effect was found in this case. In Video1, due to the prominent content of the positive film clips being a woman simulating an orgasm in front of a man in a restaurant, female subjects especially some Asian female subjects in this experiment might feel more uneasy or embarrassed than the male subjects so that they attended to evaluate the robot face and the video separately more often than the male subjects. As a result of this, the male subjects' reinterpretation of the robot expressions was more susceptible to the reinterpretation of the film clips than that of the female subjects, which might have resulted in the male subjects' average recognition accuracy of the robot expressions being lower than that of female subjects.

Therefore the preliminary experiment is mainly designed to find two pairs of film clips, one with the ability of inducing strong positive or negative moods in the subjects, and the other with the ability of inducing neutral moods to the subjects. Four positive/neutral film clips and four negative/neutral film clips will be provided for the subjects in the preliminary experiment. These film clips (see Table 6.1) will have less embarrassing content than the positive film clip used in the previous 11 experiments, and they are expected to be good choices which will be recognized well. Note that the negative film clip "The Champ's Death" listed in the table below was used before in the previous 11 experiments, and subjects recognized it as such with 100% accuracy in every experiment. It is proposed so that the recognition accuracies and the emotional valence of other negative film clips can be compared with that of this particular negative film clip. The recognition rate (measured by the fixed-choice question as used in the previous experiments), the emotional valence (measured by the Self-Assessment-Manikin scale used in the previous experiment), and feasibility to induce mood states to subjects (measured by the Affect Grid) of the new film clips must be addressed in the preliminary experiment. It will be also important to report any cross-cultural result and gender effect in the preliminary experiment.

Table 6.1 Choices of film clips and musical pieces provided in the preliminary experiment

Type of contexts	Name (time length)	Accessibility
Positive/ Neutral Film Clips	Collection of Funny Animals (3'07)	Originally edited by Schneider Yuri http://v.youku.com/v_show/id_XMjkzNTE1OTIw.html
	Lucky Louie: Why? (2'44)	Originally from American HBO TV Series Lucky Louie http://v.youku.com/v_show/id_XMjA5Nzg0NjQ0.html
	Virtual Trip of China (3'05)	Uploaded by Ms陈小西144 http://my.tv.sohu.com/u/vw/28044064
	Remember to Breathe (2'59)	Originally from Travel Video of Albert http://my.tv.sohu.com/u/vw/15587419
Negative Film Clips	The Champ's Death (2'45)	Originally from the film "The Champ" http://www.youtube.com/watch?v=FAhrqKqK_cA
	Atreyu and His Horse Artax Cross the Swamps of Sadness (2'55)	Originally from the film Neverending Story (see this on Youtube, which was uploaded by leangrysock)
	Inspiring Action to Stop Climate Change (3'00)	Produced by GREEN.tv http://www.5min.com/Video/Inspiring-Action-to-Stop-Climate-Change-517350943
	Extreme Weather Events in 2011 (3'04)	Produced by GREEN.tv http://www.5min.com/Video/2011s-Extreme-Weather-Events-517238185
Neutral Music	By a Moving River (6'08)	Originally from Ambient World by David Arkenstone http://togrool.com/download-david-arkenstone-by-a-moving-river-132076.xhtml
	Oborozukiyo (6'04)	Originally from Ambient World by David Arkenstone http://togrool.com/download-david-arkenstone-oborozukiyo-132178.xhtml
	Orinoco Dreams (4'01)	Originally from Mistyland by Bandari http://www.xiami.com/song/play?ids=/song/playlist/id/2088704/object_name/default/object_id/0
	Nocturne No.2 in E Flat Major (4'54)	Originally from The 50 Greatest Pieces of Classical Music recorded by London Philharmonic Orchestra http://www.xiami.com/song/play?ids=/song/playlist/id/1770060178/object_name/default/object_id/0

Another purpose of the preliminary experiment is the identification of a piece of music that has the ability of inducing neutral moods to subjects will be selected from the four choices as listed in Table 6.1. The feasibilities of the four choices of inducing mood states to subjects will be measured by the Affect Grid.

6.1.2 The Experimental Conditions

Subjects will be divided into 2 groups, namely, Group 1: Neutral Music and Positive/Neutral Film Clips, and Group 2: Neutral Music and Negative Film Clips. Subjects in Group 1 will first listen to two pieces of neutral music (i.e., “By a Moving River” and “Orinoco Dreams”) in a random order, and then watch four positive/neutral film clips (i.e., “Collection of Funny Animals”, “Lucky Louie: Why?”, “Virtual Trip of China”, and “Remember to Breathe”) in a random order. And subjects in Group 2 will first listen to two pieces of neutral music (i.e., “Oborozukiyo” and “Nocturne No.2 in E Flat Major”) in a random order, and then watch four negative film clips (i.e., “The Champ’s Death”, “Atreyu and His Horse Artax Cross the Swamps of Sadness”, “Inspiring Action to Stop Climate Change”, and “Extreme Weather Events in 2011”) in a random order. Note that the random order arranging of the musical pieces and the film clips will counterbalance the influence of the preceding emotional material on the subsequent emotional material.

Subjects can be recruited from the university mailing list, or leaflets distributed to the locations in the university. Subjects of both genders will be recruited and there will be no bias on the gender or the ethnic background of the subjects.

6.1.3 The Generic Experimental Procedures

In the experiment, subjects will be allocated to a warm, bright, and quiet room with only the experimenter inside. Each participant will be seated on a high back office chair, in front of a long rectangular desk. Each participant can adjust the height of the chair and distance between the chair and the desk such that he/she can view computer screen on the middle of the desk properly. During each experimental section, each participant can listen to the music and watch the film clips played by the computer, and listen to them through a stereo headphone, but he/she cannot operate the computer.

Before the experiment begins, each participant will be first asked to fill in his/her information as follows:

Male/Female:__________
Age:________
Nationality:____________________________

By doing this, gender or cross-cultural difference of the subjects on the mood effect of the music and on the recognition of the film clips can be reported. Then each of them will be asked to read the instructions for the Affect Grid and fill in one copy of it before listening to the music. He/she will fill in another two copies of it, one after the first piece of music, and the other after the second piece of music, such that the mood states of the subjects will be measured twice after listening to one piece of music. Then they will watch the four film clips in a random order, and rate each of them separately after watching it to indicate the recognition accuracy of each film clip. But before rating each film clip, they will fill in a copy of the Affect Grid. After they have watched and rated all the film clips, they will compare the emotional valence of all the four film clips by filling in the Self-Assessment-Manikin scale.

For detailed experimental procedures of the preliminary experiment, please refer to Appendix Ⅸ: The Procedures of the Preliminary Experiments.

6.1.4 The Statistical Analysis Techniques

Mean and standard deviation of the recognition accuracy and the emotional valence of each positive film clip will be compared. An independent-samples T-test in PASW 18 will be conducted to explore the impact of each piece of music on subjects' mood states, as measured by the Affect Grid twice. The gender difference and cross-cultural difference (e.g., Western subjects VS. Non-Western subjects) on the mood effect of the music and on the recognition of each film clip will be examined by Fisher's Exact Test in PASW 18. Results of the tests will provide evidence for the experimenter to select one piece of music than can induce neutral moods to subjects, and to select two pairs of film clips (one pair with strong emotional valence, and the other with neutral emotional valence) that have good recognition accuracies.

6.1.5 Predictions of the Experimental Results

After the primary data is collected in the preliminary experiment, the experimental results can be predicted listed in Table 6.2. Due to the nature of the content of the music and film clips, it is expected that no cross-cultural result or gender effect will be reported for the music or the film clips. It is expected that the four pieces of music will be observed to have the abilities of inducing neutral moods to the subjects, although they will vary in the strength of the mood they induce. For instance, they will mostly appear on the central three columns of the Affect Grid in which the leftmost column indicates the most unpleasant feelings and the rightmost column indicates the most pleasant feelings.

It is expected that subjects will recognize the two film clips "Collection of Funny Animals" and "Lucky Louie: Why?" as mostly positive, and will rate their emotional valence as positive. And the other two film clips used in Group

1 are expected to be recognized as mostly neutral, and their emotional valence will be expected to be rated as neutral or positive/neutral. As for film clips in Group 2, “The Champ’s Death”, and “Atreyu and His Horse Artax Cross the Swamps of Sadness” will be expected to be recognized as mostly negative, and their emotional valence will be expected to be rated as negative. Meanwhile, the other two film clips in Group 2 will be expected to be recognized as mostly negative, and they will be expected to be rated as negative valence.

These predictions are made after comparing the content of the film clips with that of previously investigated emotional materials, the emotional valence of which can be found in Fig. 5.1 in Chapter 5. “Collection of Funny Animals” and “Lucky Louie: Why?” can be as amusing as the positive film clip investigated previously, but they contain less embarrassing content so that it is expected that subjects will recognize them as positive and rate them as of appropriate positive valence. “Atreyu and His Horse Artax Cross the Swamps of Sadness” is expected to be recognized as negative as “The Champ’s Death”, and be rated as of negative valence as it, due to the fact that both of them share a common theme – death. And the other two film clips used in Group 1 are more or less like the previously investigated positive/neutral News recordings, while the other two film clips in Group 2 are more or less like the previously investigated negative News recordings. This is because the other two film clips used in Group 1 are likely to be neutral or weak positive, while the other two film clips in Group 2 are likely to be weak negative, judging on their content. And according to Table 4.3 in Chapter 4, the content of the recorded BBC World News is likely to be neither strong positive nor strong negative, and the emotional valences of them illustrated in Fig. 5.1 confirmed this speculation. Despite the weak emotional valences, these News recordings were recognized well by the subjects in the original experiments.

The piece of music (e.g., “By a Moving River”) that induced the most neutral moods in subjects will be selected by the experimenter. And it is expected that due to the nature of the content of the film clips, the strength of the mood they induce will vary from strong negative to strong positive. Based on different strength of the mood, two pairs of film clips will be selected by the experimenter. For example, the first pair (one film clip should be recognized as mostly positive, and the other should be recognized as mostly negative) can consist of "Collection of Funny Animals" and "The Champ's Death". In contrast, the second pair (one film clip should be recognized as mostly neutral, and the other should be recognized as mostly negative) can be "Virtual Trip of China" and "Extreme Weather Events in 2011". Consequently, a mood effect will be expected to be generated by the first pair of film clips while no mood effect is expected to be found for the second pair of film clips when conducting future new experiments.

Table 6.2 Expected results of the preliminary experiment

Group	Name	Expected Mood Effect	Expected Recognition Result	Expected Emotional Valance
Group 1	By a Moving River (Music)	Neutral	No Measurement	No Measurement
	Orinoco Dreams (Music)	Neutral	No Measurement	No Measurement
	Collection of Funny Animals	Strong Positive	Mostly positive	Positive
	Lucky Louie: Why?	Strong Positive	Mostly positive	Positive
	Virtual Trip of China	Neutral	Mostly neutral	Neutral
	Remember to Breathe	Positive/Neutral	Mostly neutral	Positive/Neutral
Group 2	Oborozukiyo (Music)	Neutral	No Measurement	No Measurement
	Nocturne No.2 in E Flat Major (Music)	Neutral	No Measurement	No Measurement
	The Champ's Death	Strong Negative	Mostly negative	Negative
	Atreyu and His Horse Artax Cross the Swamps of Sadness	Negative	Mostly negative	Negative
	Inspiring Action to Stop Climate Change	Neutral	Mostly negative	Negative
	Extreme Weather Events in 2011	Neutral	Mostly negative	Negative

If a contextual effect was found using these more carefully chosen film clip, this would suggest that the absence of a contextual effect found in the original experiments was mainly because the previous positive film clip was indeed a poor choice. And if even for the better chosen film clips no contextual effect is found, then it would suggest that obtaining contextual effects from multi-channel contexts is not feasible.

6.2. The Proposed Three New Experiments

6.2.1 The Motivation of the New Experiments

The main motivation of the three new experiments reported here is to find out whether film clips can ever work as a context on the condition of being tested in a more controlled environment. Four possible factors that may have contributed to the contextual effects found in the previous 11 experiments, namely, the choices of the contexts, the rating scheme, the mood effect, and the manner of presenting the robot face and the context, were identified in Chapter 5. When the three new experiments are conducted, these factors need to be addressed and controlled.

As the three new experiments will employ the same manner of presenting the robot face and the context as in experiment Video1 described in Chapter 4, three factors are left to be controlled. For scientific research purposes – in order to find out which factor contributes most to a contextual effect, the three new experiments will only manipulate one factor in each experiment. For instance, New Experiment One will only manipulate the choices of the contexts factor by using a new pair of film clips with the ability of inducing strong positive or negative moods to the subjects. It is proposed so because the original pair of film clips previously investigated in Video1 was observed to have the ability of inducing strong positive or negative moods in the subjects in Video2. If a contextual effect was found in New Experiment One, it can be concluded that the choices of the contexts play an important role in obtaining a contextual effect. The mood effect factor can be controlled in New Experiment Two in which another pair of film clips with the ability of inducing only neutral moods in subjects will be used. It was argued before in Chapter 4 that, a mood effect cannot be considered to be a sufficient factor for a contextual effect. If the absence of a mood effect leads to an occurrence of a contextual effect in New Experiment Two, it can be concluded that a mood effect is not necessary in obtaining a contextual effect. The reason for using a pair of film clips with the ability of inducing only neutral moods in subjects in an attempt to obtain a contextual effect is that, as mentioned earlier, this pair of film clips is expected to be more or less like the previously investigated News recordings. The News recordings were not observed to be strong valenced material as illustrated in Fig. 5.1, but yet had the ability of exerting a contextual effect in the original experiments. New Experiment Three will use which ever pair of film clips is found to exert a contextual effect in the first two new experiments and it will be made sure that only the rating scheme factor is manipulated in the third new experiment. As for the rating scheme, a more flexible rating scheme called the Geneva Emotion Wheel (see Fig. 6.1) will be used in New Experiment Three to replace the fixed rating scheme used in the previous 11 experiments.

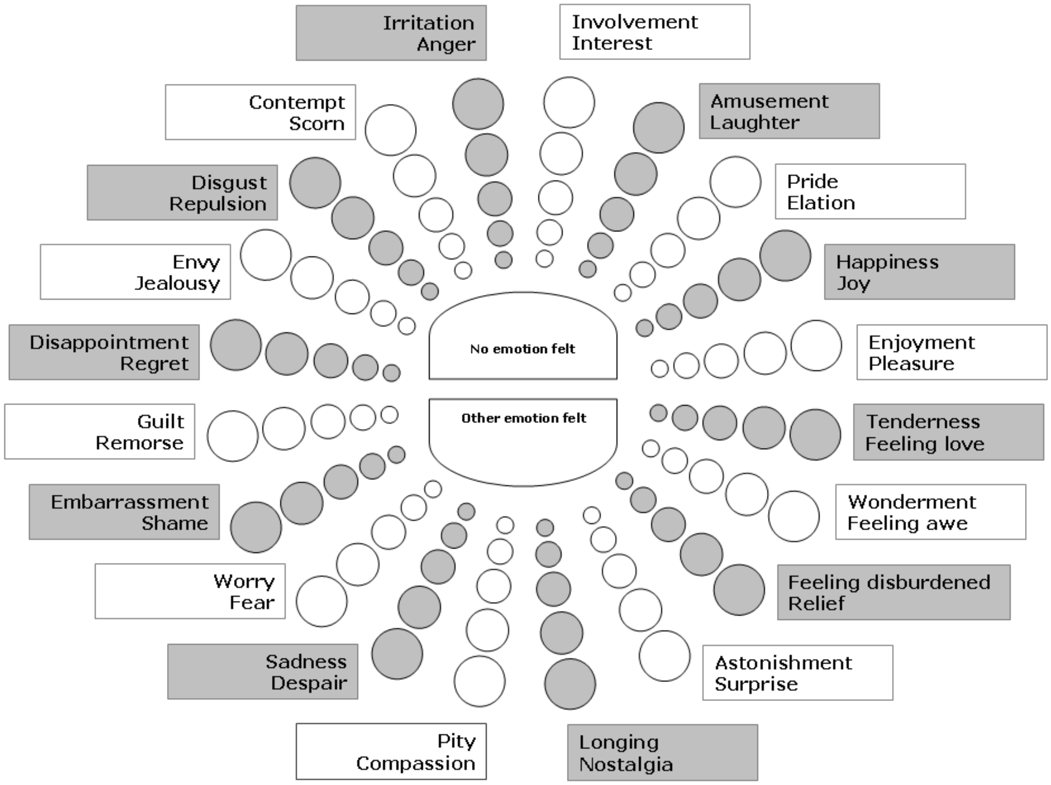


Fig. 6.1: The Geneva Emotion Wheel (Scherer, 2005) which has 20 emotion items (families) which themselves have double labels. Five scales are available for the intensity rating of each item, and "No emotion" and "Other emotion" fields increases the self-report options.

Note that before the subjects watch the film clip in each new experiment, a piece of music with the ability of inducing neutral moods to subjects will be played to the subjects to achieve a more consistent emotional state to begin with. In detail, subjects in each new experiment will first listen to a piece of music to achieve a neutral mood state, then they will fill in the Affect Grid twice, once before and once after they watch the film clip, so that it will be simple to measure the influence of the film clip on subjects' mood states. This particular piece of music, and the two new pairs of film clips as mentioned above, will be selected by the experimenter based on the results of the preliminary experiment. The idea of using music to induce moods in subjects was inspired by Niedenthal and Setterlund (1994)'s evidence of emotion congruence in perception. Although their study was designed to create strong mood effects on subjects, it did show that subjects' mood states can be controlled before the experiment begins, which might inspire a way of using calm or soothing music to neutralize any moods in subjects. In their experiments, they firstly used happy (12.5 min) or sad music (15.5 min) to induce happy or sad feelings, which could be measured by Brief Mood Introspection Scale, in subjects. Then subjects were asked to perform a lexical decision task, which consisted of 30 nonword trials and 30 word trials that contained six words from each of the five categories: happy, positive (happy unrelated), sad, negative (sad unrelated), and neutral. By analyzing the mean response latencies, Niedenthal and Setterlund found that happy subjects made lexical decisions about happy words faster than sad subjects, and sad subjects made lexical decisions about sad words faster than happy subjects. In summary, their findings showed that individuals perceive emotion-congruent information more efficiently than people experiencing a different emotional state.

6.2.2 The Hypotheses

The same two hypotheses are formulated for each of the three new experiments but not for the preliminary experiment. Earlier studies in the former chapters with the robot face have shown that the way in which facial and contextual cues were combined (congruent or incongruent with each other), affected observers' judgments about the facial expressions of the robot. Contextual effects were found for three types of contexts, namely, speech audio, still images and music. In order to examine whether such effects will also be found for video, the methodology used in these studies in which contextual effect were observed was used in selecting the way in which the facial cues of a robot and the video cues were combined. Therefore, Hypothesis 1 (H1) that subjects will recognize the sequence of robotic facial expressions better in a congruent context than in a conflicting context is once again formulated.

Also, to look at the relative weight of the video and the robot face on the emotional judgments of the robot face, hypothesis 2 (H2) that when the emotional valences of the context and the robot face are conflicting, the context will have a more dominant effect over the robot face in attributing the robot's facial expressions to its emotions, was also formulated. This hypothesis is based on the fact that speech audio and still images were observed to be more dominant over the robot face, and film clips are expected to be similarly dominant based on comparisons of the emotional valences of these three types of context illustrated in Fig. 5.1 in Chapter 5.

6.2.3 The Dependent Variable and the Independent Variable

In each new experiment, subjects' recognition of the positive/negative sequence of simulated facial expressions of a robot was the dependent variable. The two sequences of facial expressions are exactly as the same as that used in the previous experiments. Each sequence of facial expressions was about three minutes long: almost the same time length as the positive film clip or the negative film clip.

With respect to the independent variable, in each new experiment a pair of film clips will be chosen to be the surrounding context, which served as the independent variables, although it may not be the same pair of film clips in each new experiment. The first pair of film clips in the first new experiment is expected to have the ability of inducing strong positive/negative moods to the subjects, while the second pair of film clips in the second new experiment is expected to induce only neutral moods. And the third new experiment will use either pair of film clips observed to exert a contextual effect in the first two new experiments.

6.2.4 The Experimental Conditions

The experiment was mainly designed to find out if the independent variable (positive/negative surrounding film clips) had an effect on the dependent variable (subjects' recognition of the positive/negative sequence of the robotic facial expressions), as stated in H1. Therefore, in each new experiment, subjects will be divided into the same four groups, namely, Group 1: Positive/Neutral Film Clip with Positive Robotic Sequence; Group 2: Positive/Neutral Film Clip with Negative Robotic Sequence; Group 3: Negative Film Clip with Positive Robotic Sequence; and Group 4: Negative Film Clip with Negative Robotic Sequence. Among these groups, Group1 and Group 4 belong to the Congruent Condition, while Group 2 and Group 3 belong to the Incongruent Condition. According to H1, subjects in the Congruent Condition will recognize the sequence of the robotic facial expressions better than subjects in the Incongruent Condition. If H1 was validated, further investigation would establish whether H2, that in the Incongruent Condition subjects will recognize the context better than the sequence of the robotic facial expressions, was also validated. As proposed earlier, subjects in each new experiment will first listen to a piece of music which is expected to induce only neutral moods to subjects before they can watch the film clip and the robotic facial expressions simultaneously. Subjects' mood states will be measured by the Affect Grid three times – once before the music, once after the music, and the other after they have watched the film clip and the robotic facial expressions simultaneously. By doing this, it can be determined that whether subjects have achieved a consistent neutral emotional state to begin with, and whether the film clip affects the mood states of the subjects. The grouping of the subjects is illustrated in Fig. 6.2 as follows:

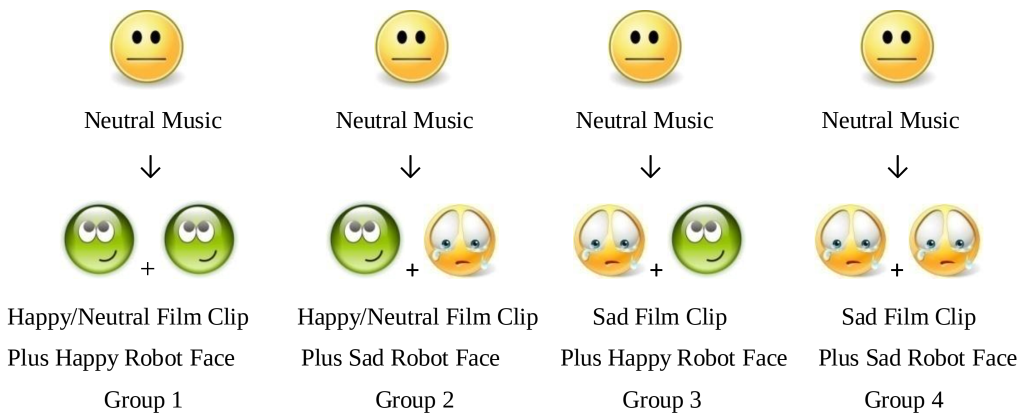


Fig. 6.2: The same grouping of each new experiment

Again, subjects can be recruited from the university mailing list, or leaflets distributed to the locations in the university. Subjects of both genders will be recruited and there will be no bias on the gender or the ethnic background of the subjects. Care will be taken to ensure that subjects in the new experiment have not taken part in the preliminary experiment.

6.2.5 The Generic Experimental Procedures

The generic experimental procedures for each new experiment are exactly the same. New Experiment Two differs from New Experiment One only in that their choices of film clips will be different, whereas New Experiment Three will use a more flexible rating scheme than that used in the first two new experiments.

Being similar to the preliminary experiment, subjects will be allocated to a warm, bright, and quiet room with only the experimenter inside. Each participant will be seated on a high back office chair in front of a long rectangular desk, and he/she can adjust the height of the chair and distance between the chair and the desk. The same robot head used in the previous experiments will be placed on the left-hand side of the desk, and the same computer screen used in the previous experiments will be placed on the right-hand side of the desk, next to the robot head. Each participant can listen to the musical piece played by the computer through a stereo headphone, and can watch a film clip played by the computer, but he/she cannot operate the computer during each experimental session.

Again, subjects will fill in their information about their gender, age, and nationalities in questionnaire as similar to Appendix Ⅲ: The Questionnaire of the Experiment Speech1. By doing this, gender or cross-cultural difference of the subjects on the recognition of the film clips or the robot expressions can be reported. At the beginning of each experiment, each participant will be asked to read the instructions for the Affect Grid and fill in a copy of it. After they have listened to a musical piece, they will fill in a second copy of the Affect Grid so that it will be measured if subjects have reached a consistent neutral mood before being showed the film clip. Then after they have watched the film clip and the robot expressions simultaneously, they will fill in a third copy of the Affect Grid, such that the mood effect of the film clip on the subjects will be measured. Finally they will answer two questions to indicate the recognition accuracy of the film clip, and of the sequence of the robot expressions.

For detailed experimental procedures of the new experiment, please refer to Appendix Ⅹ: The Questionnaire of Each Proposed New Experiment. Note that a more flexible rating scheme which will be used in the third new experiment is also provided on the last two pages of the questionnaire.

6.2.6 The Statistical Analysis Techniques

Again all null hypotheses will be rejected at the 5% level of significance, and Fisher Exact Test in PASW 18 will be chosen to test H1 and H2 in the first and the second new experiments. Due to the fact that the Geneva Emotion Wheel

(GEW) will replace the fixed choice rating scheme in the third new experiment, the T-test in PASW 18 will be chosen to test H1 and H2 instead. The interval data of each subject can be collected by summing up all the scores on the rating intensity of each item in the GEW (each of the negative 10 items in the left-hand wheel is on a scale of -5 to -1, while each of the 10 positive items in the right-hand wheel is on a scale of 1 to 5, and in the whole wheel the biggest absolute value of the scale indicates the highest intensity). By doing so, the data can indicate whether the film clip or the sequence of the robot expressions has been recognized as positive or not. For instance, if a subject chosen more positive items than the negative items in the GEW to indicate how he/she felt during watching a film clip, and the average intensity of the positive items was higher than that of the negative items, then the overall rating score in the GEW indicates that the subject recognized the film clip as positive.

As subjects' mood states will be measured three times, subjects' mood states after listening to musical piece will be tested by a T-test in PASW 18. And a two-way between-groups ANOVA in PASW 18 was chosen to analyze if the film clip or the robotic facial expressions had a mood effect on the subjects. The gender difference and cross-cultural difference (e.g., Western subjects VS. Non-Western subjects) on the recognition of the musical piece, the film clip, and the sequence of robot expressions will be examined by Fisher's Exact Test in PASW 18.

6.2.7 Predictions of the Experimental Results

Some predictions can be made on the experimental results of the three new experiments, enumerated as follows:

1. In each new experiment, before subjects watch a film clip simultaneously presented with a sequence of robot expressions, they will listen to a piece of music first. It is expected that this particular piece of

music will induce neutral moods to the subjects so that they can reach a consistent mood states before watching the video.

2. The two sequences of robot expressions used in the three new experiments will be the same as that used in previous experiments and since they were found not to color subjects' mood states, it is expected that no mood effect will be found for the robot expressions. As for the mood effect of the film clips, subjects in the first new experiment will watch the first pair of film clips chosen for their likely ability of inducing positive/negative moods, and so their mood states are expected to be affected. In contrast, subjects' mood states in the second new experiment are expected to remain neutral even after watching the film clips, as the second pair of film clips used in the experiment is chosen for their likely ability of inducing only neutral moods in subjects. And subjects' mood states in the third new experiment will be either positive/negative or neutral only, depending on which pair of the two pairs of film clips used in the first and the second new experiments will be used in the experiment. If the first pair of film clips were used in the third new experiment, then subjects' mood states after watching the video will be positive/negative, otherwise their mood states will remain as neutral.

3. The two pairs of film clips used in the new experiments are expected to achieve higher recognition accuracies than that of the film clips used in the original 11 experiments, as they are chosen more carefully. And it is expected that there will be no gender difference or cross-cultural difference (Western subjects VS. Non-Western subjects) on the perception of the film clips or the robotic expressions. Consequently, it is expected that H1 will be supported in all the three new experiments – contextual effects will be found for the two new pairs of film clips.

As mentioned earlier, the new experiments will be designed to find out if film clips can work as a context in a more controlled environment. If an effect was found in any of the new experiments, a conclusion can be arrived at that video can work as a context like the previously investigated images, speech audio, and music, and that no contextual effect was found for the previously used positive film clips mainly because they were indeed a bad choice. However, if no effect was found, even when the film clips were a better choice, it might be because combining emotional cues from multi-modality would be more difficult than from single-modality, which might reduce the contextual effect. If both a contextual effect and a mood effect were found for the first pair of film clips in the first new experiment, and only a contextual effect but not a mood effect was found for the second pair of film clips in the second experiment, this would indicate that a mood effect is not necessary to obtain a contextual effect. And if a contextual effect was not only found in the first two new experiments using the previous fixed choice rating scheme, but also in the third experiment in which a more flexible rating scheme was adopted, then a conclusion can be arrived at that a rating scheme does not play a crucial role in obtaining a contextual effect.

It is important to note that although effort will be made to minimize the effects of gender and culture on perception of the context or the robotic expressions, it is not the main purpose of conducting these new experiments. This is because gender or cross-cultural differences did not seem to play an important role in obtaining a contextual effect, as argued in Chapter 4.

6.3 Summary

This chapter proposes four new experiments, one preliminary experiment designed to identify appropriate contextual materials, and then three further experiments in which factors (i.e., the choices of the context, the mood states of the subjects, and the rating scheme) likely to affect a context effect are controlled one by one.

It is expected that by using more carefully chosen film clips, a contextual effect will be found for the film clips in more controlled environments. If a contextual effect was found for these film clips, then more experiments could be conducted to see whether such effects are only obtained under conditions of simultaneous and not separate presentation, and also to see whether a second viewing of the robot face makes a difference in the contextual effect of the film clips.

Chapter 7: Summary and Future Work

In natural settings, human receive important cues from an expresser's facial expressions and the emotion-eliciting situations in which the expression occurs. Emotions are therefore intimately related to a situation being experienced or imagined by human. The interpretation of human affect is context sensitive, for instance, human interpret a smile differently if they know that the person smiling has just heard that they have won the lottery, than they would if they knew that the same person had just been accused of bullying. Without the context, a human may misunderstand the expresser's emotional expressions. The same principle may also apply in judging a robotic facial expression in a context. Therefore it is necessary to consider the match between the robot's expressions and the surrounding contexts. This consideration is more necessary in the future as it is quite likely that human beings will encounter some socially interactive robots that have to function in the environment in unpredictable settings (hence emotional contexts).

Socially interactive robots in the future are likely to be required to function in such external environments. More specifically, in emotion based interactions, the effects of the external environment on emotional judgments of the discrete facial expressions of a robot should be looked at. Generally speaking, there are some possible factors that influence how human beings perceive a robot's expressions, such as the believability of the robotic expressions, the observers' mood states, the external environment that the observers encounter, and the manner of presenting the robot face and the context. And such factors are worth exploring. Due to the limitations of the robot head in the lab, the author of this thesis has focused more on exploring the latter three factors as enumerated above, especially the manner of presenting the robot face and the context.

7.1 Summary of the Thesis

In this thesis, it was examined how four different emotional surrounding or preceding contexts (i.e., congruent or incongruent recorded BBC News, selected affective pictures, classical music and film clips) influence users' perception of a robot's simulated emotional expressions. This has led to an understanding of how the external environment affects emotion based human-robot interactions. The robot's facial expressions were developed based on the FACS. The selection of affective pictures, musical pieces and film clips, was based on previous research showing their ability to elicit emotions, whilst the recorded pieces of BBC News were chosen from the BBC World News. These emotional materials are treated as the surrounding or preceding contexts, to simulate real-life situations to some extent.

Insights from the psychology of emotion have been taken to design the computational models of emotion for some socially interactive robots with convincing synthetic facial expressions. It has been showed that synthetic facial expressions of robots can be recognized when presented without surrounding contexts with overriding emotional valence.

Different theories of emotion not only provide insights for creating convincing synthetic facial expressions of robots, but also predict the effect of a non-neutral contextual environment on the emotional judgments about robotic facial expressions. The predictions differ depending on the theory. While Evolutionary Theories predict no contextual effect even in an emotional contextual environment, several theories, including Social Constructionist Theories, Niedenthal's theory of embodying emotion, Perceptual Control Theory (PCT), the Embodied Appraisal Theory and the Social Functionalist Approach to Emotions, are in the position to support the prediction about the situational information will influence the recognition of synthetic robot facial expressions. These predictions were used as the motivation of the thesis, which offered a conjecture that the recognition of the simulated expressions shown by a robot will be influenced by an emotional context. To explore whether surrounding or preceding emotional contexts can influence the perception and recognition of the facial emotional expressions of socially interactive robots, 11 experiments were conducted. These experiments involved a mechanical-like robot head showing either a positive sequence or a negative sequence of facial expressions and four different emotional surrounding or preceding contexts. It has been demonstrated that the interpretation of synthetic facial expressions of a robot is context sensitive, in the same way as the interpretation of human and avatar faces has been shown to be context sensitive. More specifically, the main findings are enumerated as follows:

(1) It has been examined that how four different emotional surrounding or preceding contexts (i.e., congruent or incongruent recorded BBC News, selected affective pictures, classical music and film clips) influenced users' perception of a robot's simulated emotional expressions. It turned out that three out of the four surrounding contexts have the abilities of exerting contextual effects (the particular pair of film clips used here did not). This finding suggested that the robot face was viewed in the same way as human faces, as the contextual effects found here with a robot head were also similar to the contextual effects found with human faces and avatar faces. And it indirectly supported the idea that human emotions are both biologically based and socially constructed. Therefore it supported the prediction made by certain theories of emotion that an effect of contextual environment on robotic expressions recognition will be observed, which is related to the motivation of the thesis. As a result of the contextual effects found, it was argued that that Social Constructionist Theories also apply to the recognition of the emotional expressions of a robot, and the theory of embodying emotion may also lend us support in understanding the emotional congruence observed in the recognition of the robotic expressions.

(2) It was observed that the manner in which the context and the robot head are presented (simultaneous or separate) is important in obtaining such a contextual effect. An emotional context only affected the attribution of robot emotions to its facial expressions when that context was presented simultaneously and not when it preceded the robot facial expressions.

(3) The way in which the facial cues of a robot and the contextual cues were combined (congruent or incongruent with each other), affected subjects' judgements about synthetic robot facial expressions. When there was a surrounding emotional context, people were better at recognizing the robot's facial expressions when that context was congruent with the emotional valence of its facial expressions than when the context was incongruent with the emotional valence of its facial expressions.

(4) A situational dominance was not found for every surrounding context having a contextual effect. In most of the cases, when subjects were presented with conflicting information from the robot's face and an accompanying emotional context at the same time, their attribution of robot emotions to its facial expressions were more influenced by the surrounding context than the robot face itself. However, for different surrounding contexts, there was a change in the relative weight of the surrounding context and the robot face on the emotional judgments of the robot face. Furthermore, when such effects occurred, they could be bidirectional such that the robot emotional expressions could reversely affect the recognition of the surrounding contexts.

(5) In the cases of second viewing of the robot face (observers have seen an opposite valence of robot facial expressions before the presentation of the robot face), there were still contextual effects, and no evidence that the situational dominance of the contexts was weakened.

(6) Gender or cross-cultural differences did not play an important role in obtaining a contextual effect, as for the cross-cultural differences found in two cases, a contextual effect was still observed in one case while no contextual effect was found in the other case.

(7) Four possible explanations for the presence or absence of contextual effects in these experiments were proposed, namely, the emotional embodiment explanation, the mood effect explanation, the 'robot's interpretation of the context' explanation, and the cue integration explanation. And it was suggested that the cue integration explanation may be better supported by the data (e.g., mean and standard deviation of recognition accuracy of each emotional context in the 11 experiments) than other three explanations. This is because, first of all,

no data was collected to validate the embodied recognition of robot facial expressions; secondly, there was a limited possibility that subjects interpreted the robot's expressions as if the robot was capable of "seeing" or "listening to" the context; thirdly, it was argued that a mood effect was not a sufficient factor for the occurrence of a contextual effect, although it remains uncertain whether or not an effect on subjects' mood is necessary for a context effect to occur.

(8) Three possible explanations for the absence of a contextual effect for the particular pair of film clips were proposed, namely, the lack of subjects' attention to the robot head, the difficulty in combining emotional cues from multi-modality such as video, and the questionable choice of the positive film clip. Although no hard evidence was collected to clearly indicate that the positive film clip was a poor choice, the first explanation is considered unlikely because enough care was taken by the experimenter to make sure that subjects pay enough attention to the robot head. Future experiments could be conducted to distinguish between the second and the third explanations. For instance, a better chosen pair of film clips could be used, and it could be concluded that a poor choice of film clips led to the absence of a contextual effect for the particular pair of film clips used before if a contextual effect is found for the new film clips.

To understand better the contextual effects found and not found in the 11 experiments, a detailed comparison of the recognition rate and the emotional valence of the robot's simulated emotions and each context was given, using the data extracted from a further experiment conducted after the 11 experiments. The results of the experiment reinforced the claims about the robot's expressions, which are developed based on the FACS, and supported the selection of the recorded BBC News, affective pictures, musical pieces and negative film clip. However, the experiment provided evidence that the positive film clip used in the previous experiments was indeed a poor choice, compared to the negative film clip. In order to find out why a contextual effect was not always found for each type of the contexts in all situations, four possible factors contributing to a contextual effect were identified, namely, the choices of the contexts, the mood effect, the rating scheme, and the manner of presenting the robot face and the context. It was suggested that by using a better choice of film clips and a more controlled methodology, future experiments could be conducted to explore the possibility that a video context might also be shown to exert a contextual effect under some circumstances.

In order to find out whether a contextual effect can be exerted by a video context, four new experiments were proposed. These four experiments include one preliminary one to select a piece of neutral music and two pairs of film clips of good choices, and the other three to establish whether video can ever work as a context in a more controlled environment in which the choices of the context, the mood states of the subjects, and the rating scheme are controlled one by one. Reasons for expecting that a contextual effect will be found for the film clips in a more controlled environment were discussed.

7.2 Future Work

In the future, my intention is to carry out further investigations that would lead the readers to understand the contextual effects better and coordinate the robot to respond to the external environment (i.e., the surrounding contexts) more appropriately. Further investigations into the following research areas are proposed:

(1) **Deeper probing of the effects found and not found for the emotional materials selected in the study.** It has been observed in my research that not all the surrounding contexts have contextual effects (the particular pair of film clips used here did not). Future research (including the four new experiments proposed in Chapter 6) could choose different film clips in order to examine whether other film clips might have an effect on the interpretation of the same robot face. Also, in order to see if there will be a change of the relative weight of the surrounding context and the robot face on the emotional judgments of the robot face, the contents of the surrounding contexts having effects on the robot face (i.e., the recorded BBC News, selected affective pictures, and the musical

pieces) could be modified to some extent. For example, the affective pictures used in the 11 experiments and a further experiment conducted after the 11 experiments were recognized well by the subjects and were validated to be of strong emotional valence. And the contextual effects found on the affective pictures were dominant. Future experiments could replace some of the slides of the affective pictures by some slides of weaker emotional valence to see if the contextual effects of the new set of affective pictures will become less dominant.

(2) **Examining the effects of the same emotional materials on a different robot.** The robot head used in my study has an admittedly limited ability to articulate expressions. It might be that contextual effects are found here precisely because the robot's expressions are somewhat ambiguous. Therefore, it would be interesting to know whether some robots, such as Kismet, Probo, or BERT2 Head (Bazo et al. 2010), with more expressive faces are similarly susceptible to the effects of a surrounding context or not. If effects of the same emotional materials are also found for other robots, the findings obtained in this study will have been shown to generalize.

(3) **Involving more expressive skills to look at the effects of the same emotional materials on the same or a different robot.** Note that in my experiments, the robot face is the only channel used to express emotions. Future research could look at what happens when a robot is equipped with more channels to express emotions (e.g., vocal skills or gesture skills), and whether there will be a change in the relative weight of the surrounding context and the robot face on the emotional judgments of the robot face. Note that the emotional speech or gestures should be congruent with the emotional valence of the sequence of a robot's facial expressions, so that the ambiguity of the robot face could be reduced to some extent. According to (Bazo et al. 2010), humans do not simply make static faces to express emotion, because body language, context, and time varying expressions all play a significant role in the meaning of a particular facial expression. Therefore, it can be expected that congruent speech or gestures will make a robot's facial expressions look more convincing.

(4) **Coordinating the robot to respond to the surrounding contexts appropriately.** As the results of this thesis demonstrate, one way of making a

robot’s emotional expressions seem more convincing, is to make sure that they are supported by, and match, the surrounding emotional context. Yet no empirical research has looked into the issue of coordinating the robot to respond to the surrounding emotional contexts appropriately. Hong et al. (2002) used neural networks to map real-time surrounding emotional contexts to synthetic avatar emotions – similar approaches in which the valence of surrounding context is detected, and the robot’s expressions adapted to match it, may well prove to be a good way of creating robots in the future that are seen as more convincing and believable. And another way of coordinating the robot to respond to the surrounding emotional contexts appropriately would be adopting Moore’s (2007) PCT based model called PRESENCE “PREdictive SENsorimtor Control and Emulation”, as mentioned in Chapter 2. Moore’s model could enable a system to have in mind the needs and attention of a user while a user has in mind the needs and intentions of the system. According to Moore (2007), cooperative and communicative behavior emerges as a by-product of the recursive hierarchical feedback control structure that is founded on this model of integration. Therefore, PRESENCE suggests a new model of robotic expressions recognition that can enable a robot to selects its facial expressions appropriate to the needs and attention of an observer by taking into account the robot's emulation of the observer's emulation of the robot’s intentions and motor behavior (facial expressions). The observer's emulation of the robot’s intentions relies on the observer’s knowledge of whether the robot’s emotional expressions will be supported by, and match, the surrounding emotional context. As a result of this, the observer's emulation of the robot’s motor behavior should take into account the robot’s capability of coordinating its facial expressions influenced by the surrounding contexts to communicate with the observer. “Sensitivity” in the four-layer architecture of PRESENCE can represent the degree to which the robot takes into account the disturbance of the contexts. In other words, the robot must evaluate the influence of the surrounding contexts on the observer’s recognition of its facial expressions before coordinating them to respond to the observer.

References

Alexandersson, J., Becker, T., Engel, R., L ̈ockelt, M., Pecourt, E., Poller, P., Pfleger, N., & Reithinger, N.: Ends-based dialogue processing, In Proceedings Second International Workshop on Scalable Natural Language Understanding, Boston. (2004)

Alonso, J.: Studying social cues in human robot interaction, Microsoft External Research Symposium, MIT Media Laboratory, Personal Robots. (2009)

Armon-Jones, C.: The thesis of constructionism. In The Social Construction of Emotions, ed. R. Harr é, 2–56. Oxford: Blackwell. (1986)

Arnold, M. B.: Emotion and personality: vol. 1. Psychological aspects. New York, NY: Columbia University Press. (1960)

Aronson, E.: Back to the Future: Retrospective Review of Leon Festinger's "A Theory of Cognitive Dissonance". The American Journal of Psychology, 110, 1, 127-137. (1997)

Aucouturier, J.-J., Ogai, Y. & Ikegami, T.: Using Chaos to Trade Synchronization and Autonomy in a Dancing Robot, IEEE Intelligent Systems, vol.23, No.2. (2008)

Averill, J. R.: A constructivist view of emotion. In R. Plutchik and H. Kellerman (eds.), Emotion: Theory, research and experience: Vol. I. Theories of emotion (pp. 305-339). New York, NY: Academic Press. (1980)

Baenziger, T., Tran, V. and Scherer, K. R.: ‘‘The Emotion Wheel. A Tool for the Verbal Report of Emotional Reactions’’, poster presented at the conference of the International Society of Research on Emotion, Bari, Italy. (2005)

Bailey, A.: Synthetic Social Interactions with a Robot using the BASIC personality model. MSc thesis, University of Sheffield, Sheffield, England (2006)

Barrett, L. F.: Are emotions natural kinds? Persp Psychol Sci1:28–5. (2006)

Barsalou, L.W.: Situated simulation in the human conceptual system, Lang. Cogn. Process, 18, 513–562. (2003)

Barsalou, L. W., Niedenthal, P. M., Barbey, A., & Ruppert, J.: in The Psychology of Learning and Motivation, B. H. Ross, Ed. (Academic, San Diego, CA, 2003), vol. 43, pp. 43–92. (2003)

Bartels, A. and Zeki, S.: The neural basis of romantic love. Neuroreport, 11, 829-834. (2000)

Bartlett, M. S.: Emotion simulation and expression understanding: A case for time. Invited commentary on Niedenthal, Mermillold, Meringer, and Hess, The Simulation of Smiles (SIMS) model: Embodied simulation and the meaning of facial expression. Behavioral and Brain Sciences 33(6) p. 435-436. (2010)

Bazo D., Vaidyanathan, R., Lenz, A., & Melhuish, C.: Design and Testing of Hybrid Expressive Face for the BERT2 Humanoid Robot, IEEE International Conference on Intelligent Robots and Systems (IROS), pp 5317-5322, Taipei, Taiwan, Oct. (2010)

Beale, R., & Creed, C.: Affective interaction: How emotional agents affect users, Int. J. Human-Computer Studies 67, 755–776. (2009)

Becker-Asano, C. & Ishiguro, H.: Evaluating facial displays of emotion for the android robot Geminoid F. IEEE SSCI Workshop on Affective Computational Intelligence, pp. 22-29. (2011)

Bell, C.: Essays on the Anatomy of Expression in Painting, London: Longman, Hurst, Rees & Orme, pp.127–9. (1806)

Berry, D., Butler, L. T., & DeRosis, F.: Evaluating a realistic agent in an advice-giving task. International Journal of Human–Computer Studies 63(3), 304–327. (2005)

Bradley, M. M., & Lang, P. J.: The International Affective Picture System (IAPS) in the Study of Emotion and Attention. In J.A. Coan & J.B. Allen (Eds.), Handbook of Emotion Elicitation and Assessment (pp. 29–46). New York: Oxford University Press. (2007)

Breazeal, C.: Early experiments using motivations to regulate human-robot interaction, Proc. 1998 Amer. Assoc. Artif. Intell. (AAAI98) Fall Symposium on Emotional and Intelligent: the Tangled knot of Cognition, Techn. Rep., Vol. FS-98–03, (Orlando) pp. 31–36. (1998)

Breazeal, C.: Emotive qualities in robot speech, IEEE International Conference on Intelligent Robots and Systems 3, pp. 1388-1394. (2001)

Breazeal, C.: Designing sociable robots. MIT Press, Cambridge. (2002)

Breazeal, C.: Emotion and sociable humanoid robots, Int. J. Human Comput. Interact. 59, 119–155. (2003a)

Breazeal, C.: Toward sociable robots, Robotics and Autonomous Systems, 42(3-4): pp. 167-175. (2003b)

Breazeal, C., & Aryananda, L.: Recognizing affective intent in robot directed speech, Auton. Robots 12(1), 85–104. (2002)

Breazeal, C., Brooks, A., Gray, J., Hoffman, G., Kidd, C., Lee, H., Lieberman, J., Lockerd, A., & Chilongo, D.: Tutelage and collaboration for humanoid robots, Int. J.Humano. Robots 1(2), 315–348. (2004)

Breazeal, C., & Scassellati, B.: Infant-like social interactionsbetween a robot and a human caregiver,Adaptive Behavior 8(1), 47–72. (2000)

Breazeal, C., Takanishi, A., & Kobayashi, T.: Social Robots that Interact with People. Springer Handbook of Robotics, pp.1349~1369. (2008)

Brooks, A., Gray, J., Hoffman, G., Lockerd (Thomaz), A., Lee, H., & Breazeal, C.: Robot's play: interactive games with sociable machines, Computers in Entertainment, 2(3):1–18, July. (2004)

Burleson, W., & Picard, R.: Gender-specific approaches to developing emotionally intelligent learning companions, Intelligent Systems 22 (4), 62–69. (2007)

Calinon, S. & Billard, A.: Teaching a Humanoid Robot to Recognize and Reproduce Social Cues, IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). (2006)

Canamero, L. & Fredslund, J.: How does it feel? Emotional interaction with a humanoid lego robot. Socially intelligent agents: the human in the loop. Papers from the AAAI 2000 fall symposium, pp 23–28. (2000)

Carrera-Levillain, P., Fernandez-Dols, J.-M.: Neutral faces in context: Their emotional meaning and their function. J Nonverb Behav, 18:281-99. (1994)

Carroll, J. M., & Russell, J. A.: Do facial expressions signal specific emotions? Judging the face in context. J. Personal. Soc. Psychol. 70: 205–18. (1996)

Cauldwell, R. T.: Where did the anger go? The role of context in interpreting emotion in speech. In Speech and Emotion: Proceedings of the ISCA workshop, Newcastle, Co. Down, N. Ireland, September 2000 (eds R. Cowie, E. Douglas-Cowie & M. Schroeder), pp. 127–131. (2000)

Chartrand, T. L., & Bargh, J. A.: The Chameleon Effect: The Perception-Behavior Link and Social Interaction, Journal of Personality and Social Psychology, 76, 893-910. (1999)

Clore, G. L. & Ortony, A.: Cognition in emotion: Always, sometimes, or never? In L. Nadel, R. Lane & G. L. Ahern (Eds). The Cognitive neuroscience of emotion. New York: Oxford University Press. (2000)

Coan, J., & Allen, J.: Organizing the Tools and Methods of Affective Science, Introduction in J.A. Coan & J.J.B Allen (Eds.), Handbook of Emotion Elicitation and Assessment (pp. 3-6). New York: Oxford University Press. (2007)

Coan, J. A., & Gottman, J. M.: The specific affect (SPAFF) coding system. In J. A. Coan and J. J. B. Allen (Eds.) Handbook of Emotion Elicitation and Assessment (pp. 267-285), New York, NY: Oxford University Press. (2007)

Cohn, J. F.: Advances in behavioral science using automated facial image analysis and synthesis. IEEE Social Signal Processing Magazine, 128(6), 128-133. (2010)

Cohn, J. F., & Sayette, M. A.: Spontaneous facial expression in a small group can be automatically measured: An initial demonstration. Behavior Research Methods, 42(4), 1079-1086. (2010)

Cole, J.: About Face. Cambridge, MA: The MIT Press. (1998)

Cosmides, L. & Tooby, J.: From evolution to behavior: Evolutionary psychology as the missing link. In J.

Dupre (Ed.), The latest on the best: Essays on evolution and optimality. Cambridge, MA: The MIT Press. (1987)

Cosmides, L. & Tooby, J.: Evolutionary psychology and the emotions In M. Lewis & J. M. Haviland-Jones (Eds.), Handbook of Emotions, 2nd Edition. (pp. 91-115.) NY: Guilford. (2000)

Cowie, R.: Perception of emotion: towards a realistic understanding of the task. Phil. Trans. R. Soc. B (2009) 364, 3515-3525. (2009)

Creed, C., & Beale, R.: Using Emotion Simulation to Influence User Attitudes and Behaviour, to be presented at HCI 2005, Edingburgh, September. (2005)

Creed, C. & Beale, R.: Psychological Responses to Simulated Displays of Mismatched Emotional Expressions. Interacting with Computers, 20(2), pp. 225-239. (2007)

Damasio, A.R.: Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition, Cognition, 33: 25-62, 1989. (1989)

Damasio, A. R.: Descartes' Error: Emotion, Reason, and the Human Brain (Grosset/Putnam, New York, 1994). (1994)

Damasio, A. R.: The feeling of what happens: Body and emotion in the making of consciousness. New York, NY: Harcourt Brace & Company. (1999)

Darwin, C.: The expression of the emotions in man and animals**.** London: John Murray. (1872)**;** 3rd edn (ed. P. Ekman). London: Harper Collins; New York: Oxford University Press. (1998)

Dautenhahn, K., Bond, A. H., Canamero, L., & Edmonds, B.: Socially Intelligent Agents: Creating Relationships with Computers and Robots, Multi agent Syst. Artif. Soc. Simul. Organiz., Vol. 3, Springer, Berlin, Heidelber. (2002)

Dautenhahn , K., & Werry, I.: Issues of robot–human interaction dynamics in the rehabilitation of children with autism, in Proc. From Animals to Animats, Sixth Int. Conf. Simul. Adaptive Behavior, Cambridge, MA, Sept. 11–15, pp. 519–528. (2000)

Decety, J., & Jackson, P. L.: The Functional Architecture of Human Empathy, Behav Cogn Neurosci Rev, June 1, 2004; 3(2): 71-100. (2004)

De Gelder, B., & Vroomen, J.: The perception of emotions by ear and by eye. Cognition & Emotion, 14(2000):289–311. (2000)

Dewey, J.: The reflex arc concept in psychology. Psychological Review, 3, 357–370. (1896)

Duchenne, G. B. In: The Mechanism of Human Facial Expression. Cuthbertson RA, translator and editor. New York: Cambridge University Press; (1990) (Reprinted from Mécanisme de la Physionomie Humaine, 1862, Paris, Jules Renouard Libraire)

Eich, E., Ng, J.T.W., Macaulay, D., Percy, A.D., & Grebneva, I.: Combining music with thought to change mood. In J.A. Coan & J.B. Allen (Eds.), Handbook of Emotion Elicitation and Assessment (pp.124-136). New York: Oxford University Press. (2007)

Eisenberg, N., Fabes, R.A., Miller, P.A., Fultz, J., Shell, R., Mathy, R.M., & Reno, R.R.: Relation of sympathy and distress to prosocial behavior: A multi-method study. Journal of Personality and Social Psychology, 57, 55-66. (1989)

Ekman, P., Sorenson, E. R. & Friesen, W. V.: Pan-cultural elements in facial displays of emotions. Science, 164, 86-88. (1969)

Ekman, P.: Universals and cultural differences in facial expressions of emotion. In J. K. Cole (Ed.), Nebraska Symposium on Motivation (Vol. 19, pp. 207-283). Lincoln: University of Nebraska Press. (1972)

Ekman, P.: Movements with precise meanings.J. Commun. 26, 14–26. (1976)

Ekman, P.: Telling lies. New York, NY: WW Norton and Company. (1985, 2009)

Ekman, P., et al.: Universals and Cultural Differences in the Judgments of Facial Expressions of Emotion. Journal of Personality and Social Psychology, 53(4), 712-717. (1987)

Ekman, P.: Basic emotions. In T. Dalgleish and T. Power (Eds.) The Handbook of Cognition and Emotion. pp. 45-60. (1999)

Ekman, P.: Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life, Henry Holt & Co. (2004)

Ekman, P.: Darwin's contributions to our understanding of emotional expressions. Phil. Trans. R. Soc. B 364, 3449–3451. (2009)

Ekman, P., Friesen,W.: Constants Across Cultures In The Face And Emotion. Journal of Personality and Social Psychology, 1971, 17, 124-129. (1971)

Ekman, P., Friesen, W.: Facial Action Coding System. Consulting Psychologists Press (1978)

Ekman, P., Friesen, W., Hager, J.: Facial Action Coding System. Research Nexus, Salt Lake City, Utah (2002)

Elfenbein, H. A., & Ambady, N.: On the universality and cultural specificity of emotion recognition: a meta-analysis. Psychol. Bull. 128:203–35. (2002)

Faigin, G.: The Artist’s Complete Guide to Facial Expression, Watson Guptill, New York, NY. (1990)

Fairchild, G., Van Goozen, S., Calder, A., Stollery, S., & Goodyer, I.: Deficits in facial expression recognition in male adolescents with early-onset or adolescence-onset conduct disorder. Biol Psychiatry. (2009)

Fernández-Dols, J. M., Wallbott, H., & Sanchez, F.: Emotion Category Accessibility and the Decoding of Emotion from Facial Expression and Context. Journal of Nonverbal Behavior, 15, 107-123. (1991)

Fernández-Dols, J. M.; Carroll, J. M.: Is the meaning perceived in facial expression independent of its context? In J.A. Russell and J.M. Fernández-Dols (Eds.) The psychology of facial expression Cambridge: Cambridge University Press. (1997)

Fogg, B.: Prominence-Interpretation Theory: Explaining How People Access Credibility Online, (Eds.), CHI '03 extended abstracts on Human factors in computing systems, Florida, USA, pp.722 – 723. (2003)

Fong, T., Nourbakhsh, I., & Dautenhahn, K.: A Survey of Socially Interactive Robots: Concepts, Design, and Applications, Technical Report CMU-RI-TR-02-29. (2002)

Frijda, N. H. & Scherer, K. R.: Emotion Definition (psychological perspectives). In Oxford companion to emotion and the affective sciences (eds D. Sander & K. R. Scherer), pp. 142–143. Oxford, UK: Oxford University Press. (2009)

Fujita, M.: On activating human communications with pet-type robot AIBO, Proc. IEEE 92(11), 1804–1813. (2004)

Gockley, R., Simmons, R., & Forlizzi, J.: Modeling affect in socially interactive robots, Proc. 15th IEEE Int. Symp. Robot Human Interact. Commun. pp. 558–563. (2006)

Gorostiza, J.F., Barber, R., Khamis, A. M., Malfaz, M., Pacheco, R., Rivas, R., Corrales, A., Delgado, E., & Salichs, M.A.: Multimodal Human-Robot Interaction Framework for a Personal Robot, RO-MAN 06: The 15th IEEE International Symposium on Robot and Human Interactive Communication, Hatfield, United Kingdom, Sep. (2006)

Gray, J.A.: Elements of a Two-Process Theory of Learning. London: Academic Press. (1975)

Gray, J.A.: The Psychology of Fear and Stress, 2nd edn. Cambridge: Cambridge University Press. (1987)

Green, J. & D'Oliveira, M.: Learning to use statistical tests in psychology. Buckingham, UK: Open University Press. (1999)

Gunes, H., Piccardi, M., & Pantic, M.: From the Lab to the real world: affect recognition using multiple cues and modalities. In: Affective computing: focus on emotion expression, synthesis, and recognition, In Tech Education and Publishing, Vienna, Austria, pp.185-218, ISBN 978-3-902613-23-3. (2008)

Haidt, J.: The moral emotions. In R. J. Davidson, K. R. Scherer & H. H. Goldsmith (Eds.), Handbook of affective sciences (pp. 852-870). Oxford, UK: Oxford University Press. (2003)

Haidt, J., & Keltner, D.: Culture and facial expression: Open-ended methods find more expressions and a gradient of recognition. Cogn. Emot. 13:225–66. (1999)

Hashimoto, M., Hashizume, H., & Katoh, Y.: Design of Dynamics for Synchronization Based Control of

Human-Robot Interaction, Proceedings of the 2006 IEEE International Conference on Robotics and Biomimetics, December 17 - 20, Kunming, China. (2006)

Harmon-Jones, E., Amodio, D. M., & Zinner, L. R.: Social psychological methods of emotion elicitation. In J. A. Coan & J. J. B. Allen (Eds.), Handbook of Emotion Elicitation and Assessment, (pp. 91-105). New York: Oxford University Press. (2007)

Hayashi, K., Onishi, Y., Itoh, K., Miwa, H., & Takanishi, A.: Development and evaluation of face robot to express various face shape, Proc. IEEE Int. Conf. Robot. Autom, pp. 481–486. (2006)

Heerink, M., Kröse, B. J. A., Wielinga, B. J., & Evers, V.: The Influence of a Robot's Social Abilities on Acceptance by Elderly Users, In Proceedings RO-MAN, Hertfordshire. (2006)

Held R., Ostrovsky Y., Degelder B., Gandhi T., Ganesh S., Mathur U., & Sinha P.: *The newly sighted fail to match seen with felt*. Nature Neuroscience, 14:551–553. (2011)

Henrik, C., Jaeckel, P., Campbell, N., Lawrence, N. & Melhuish, C.: Shared Gaussian Process Latent Variable Models for Handling Ambiguous Facial Expressions. In: Proceedings Intelligent Systems and Automation: 2nd Mediterranean Conference on Intelligent Systems and Automation (CISA'09), 5 March 2009. 1107, pp 147-153. (2009)

Hong, P., Wen, Z., and Huang, T.: Real-time speech driven expressive synthetic talking faces using neural networks. IEEE Transaction on Neural Networks, April. (2002)

Iida, F., Tabata, M., Hara, F.: Generating personality character in a Face Robot through interaction with human, Proc. 7th IEEE Int. Workshop Robot Human Commun. pp. 481–486. (1998)

Ikeuchi, K., Shiratori, T., Kudoh, S., Hirukawa, H., Nakaoka, S., & Kanehiro, F: Robots that Learn to Dance from Observation, IEEE Intelligent Systems, Vol. 23, No. 2, pp. 74-76. (2008)

Ishiguro, H.: Android Science: Toward a new cross-interdisciplinary framework. Proc. of the CogSci 2005 Workshop Toward Social Mechanisms of Android Science. Stresa, Italy. (2005)

Ishiguro, H.: Android Science: conscious and subconscious recognition. Connection Science, Vol. 18, No.4, 319-332. (2006)

Izard, C. E.: The face of emotion. New York: Appleton-Century-Crofts. (1971)

Jaeckel, P., Campbell, N., & Melhuish, C.: Gaussian process regression for facial behavior mapping. In Proceedings Taros 2008. (2008a)

Jaeckel, P., Campbell, N., & Melhuish, C.: Facial behavior mapping - from video footage to a robot head. Robotics and Autonomous Systems. (2008b)

James, W.: What is an Emotion? Mind, 9(34), 188-205. (1884)

Johnson-Laird, P.N., & Oatley, K.: Basic emotions, rationality, and folk theory. Cognition and Emotion, 6, 201-223. (1992)

Kahn, P. H. Jr., Freier, N. G., Kanda, T., Ishiguro, H., Ruckert, J. H., Severson, R. L., & Kane, S. K.: Design Patterns for Sociality in Human-Robot Interaction, HRI'08, Amsterdam, Netherlands, March 12–15. (2008)

Kanda, T., Miyashita, T., Osada, T., Haikawa, Y., & Ishiguro, H.: Analysis of Humanoid Appearances in Human-Robot Interaction, IEEE TRANSACTIONS ON ROBOTICS, VOL. 24, NO.3, JUNE. (2008)

Kappas, A.: The sociality of appraisals: Impact of social situations on the evaluation of emotion antecedent events and physiological and expressive reactions. In NH Frijda (Ed.), ISRE1996. (1996)

Keltner, D., Haidt, J., & Shiota, L.: Social Functionalism and the Evolution of Emotions. In M. Schaller, D. Kenrick, & J. Simpson (Eds.) Evolution and Social Psychology, pp. 115-142. (2006)

Kidd, C. D., & Breazeal, C.: Robots at Home: Understanding Long-Term Human-Robot Interaction, 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, Acropolis Convention Center, Nice, France, Sept, 22-26. (2008)

Kose-Bagci, H., Dautenhahn, K., & Nehaniv, C. L.: Emergent Dynamics of Turn-Taking Interaction in Drumming Games with a Humanoid Robot, Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, Technische Universität München, Munich, Germany, August 1-3. (2008)

Kosuge, K., Takeda, T., Hirata, Y.: Dancing with a robot: The human leads, IEEE Intelligent Systems, Vol.23, No.2, pp.82-84. (2008)

Kozima, H.: An anthropologist in the children's world: A field study of children's everyday interaction with an interactive robot, Proc. Int. Conf. Dev.Learn. (ICDL), Bloomington. (2006)

Krohne, H. W.: Stress and coping theories. In N. J. Smelser & P. B. Baltes (Eds.), The international encyclopedia of the social and behavioral sciences (Vol. 22, pp. 15163-15170). Oxford, UK: Elsevier. (2001)

Lang, P.J.: Behavioral treatment and bio-behavioral assessment: Computer applications.In J.B. Sidowski, J.H. Johnson, & T.A. Williams (Eds.). Technology in mental health care delivery systems (pp. 119–l37). Norwood, NJ, USA: Ablex. (1980)

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

Lange, C. G.: Om sindsbevaegelser: et psyko-fysiologisk studie.Kjbenhavn: Jacob Lunds. (1885) Reprinted in The emotions, C. G. Lange and W.James (eds.), I. A. Haupt (trans.) Baltimore, Williams & Wilkins Company1922.

Lazarus, R. S.: Psychological stress and the coping process. New York, NY: McGraw Hill. (1966)

Lazarus, R. S.: Emotion and adaptation. New York, NY: Oxford University Press. (1991)

Lee, M. C. T., Ng, H. H. E., Tang, S. W., & Chan, C. H. C.: Effects of sad mood on facial emotion recognition in Chinese people, Psychiatry Research 159 (2008) 37–43. (2008)

Leite, I., Pereira, A., Martinho, C., & Paiva, A.: Are Emotional Robots More Fun to Play With? Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, TechnischeUniversität München, Munich, Germany, August 1-3. (2008)

Levenson, R. W.: Autonomic nervous system differences among emotions. Psychological Science, 3, 23-27. (1992)

Levenson, R. W.: Emotion elicitation with neurological patients. In J. A. Coan and J. J. B. Allen (Eds.) Handbook of Emotion Elicitation and Assessment (pp. 158-170), New York, NY: Oxford University Press. (2007)

Leppänen, J. M. & Hietanen, J. K.: Affect and face perception: Odors modulate the recognition advantage of happy faces. Emotion, 3, 315-326. (2003)

MacLean, P. D.: Cerebral evolution of emotion. In M. Lewis and J. M.Haviland (eds.), Handbook of emotions (pp. 67-83). New York, NY: Guilford Press. (1993)

Maldonado, H., Lee, J. R., Brave, S., Nass, C., Nakajima, H., Yamada, R., Iwamura, K., & Morishima, Y.: We learn better together: enhancing eLearning with emotional characters. In: Koschmann, T., Suthers, D., Chan, T.W.(Eds.), Computer Supported Collaborative Learning 2005:TheNext10Years!. Lawrence Erlbaum Associates, Mahwah, NJ, pp.408–417. (2005)

Mayer, J. D., & Gaschke, Y. N.: The experience and meta-experience of mood. *Journal of Personality and Social Psychology, 55, pp.*102-111. (1988)

Marsella, S., Gratch, J., & Petta, P.: Computational Models of Emotion. In in Scherer, K.R., Bänziger, T., & Roesch, E. (Eds.) A blueprint for a affective computing: A sourcebook and manual. Oxford: Oxford University Press, (2010)

Mehrabian, A.: An analysis of personality theories. Prentice-Hall, Englewood Cliffs, New Jersey. (1968)

Michalowski, M. P., Kozima, H., Nichols, J., & Cornwell, M.: Rhythm in Human-Robot Social Interaction, IEEE Intelligent Systems 23:2 March/ April. (2008)

Michalowski, M. P., Sabanovic, S., & Kozima, H: A dancing robot for rhythmic social interaction, Proceedings of the 2nd Annual Conference on Human-Robot Interaction (HRI 2007), March, pp. 89 - 96. (2007)

Millenson, J.R.: Principles of Behavioral Analysis. New York:Macmillan. (1967)

Miwa, H., Itoh, K., Takanobu, H., & Takanishi, A.: Design and control of 9-DoFs emotion expression humanoid arm, 2004 IEEE Int. Conf. Robot. Autom. (New Orleans) pp. 128–133. (2004)

Miwa, H., Itoh, K., Takanobu, H., & Takanishi, A.: Mechanical design and motion control of emotion expression humanoid robot WE-4R, 15th CISM-IFToMM Symp. Robot Des. Dyn. Cont. ROM04–14. (2004)

Mobahi, H.: Building an interactive robot from scratch. Bachelor thesis. (2003)

Mori, M.: Bukimi no tani the uncanny valley, Energy 7(4), 33–35. (1970)

Mori, M.: The Buddha in the robot. Tuttle Publishing. (1981)

Moore, R. K.: PRESENCE: A human-inspired architecture for speech-based human-machine interaction. IEEE Trans. Computers, 56(9), 1176-1188. (2007)

Moors, A.: Theories of emotion causation: a review. Cogn. Emotion 23, 625–662. (2009)

Morkes, J., Kernal, H. & Nass, C.: Humor in task-oriented computer-mediated communication and human–computer interaction. In: Proceedings of International Conference on Human factors in Computing Systems (CHI98). ACM, LosAngeles,USA,pp.215–216. (1998)

Moshkina, L., & Arkin, R. C.: Human perspective on affective robotic behavior: A longitudinal study, In 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, (IROS 2005), pages 2443—2450. (2005)

Mower, E., Lee, S., Matariric, M.J., & Narayanan, S.: Human perception of synthetic character emotions in the presence of conflicting and congruent vocal and facial expressions. In IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP 2008), Las Vegas, NV, pp. 2201–2204 (2008)

Mower, E., Matarić, M. J., & Narayanan, S.: Human perception of audio-visual synthetic character emotion expression in the presence of ambiguous and conflicting information. IEEE Transactions on Multimedia, VOL.11, NO.5. (2009)

Mutlu, B., Osman, S., Forlizzi, J., Hodgins, J., & Kiesler, S.: Task Structure and User Attributes as Elements of Human-Robot Interaction Design, in Proc. IEEE RO-MAN 06 Conference, Hatfield, United Kingdom. (2006)

Naab, P. J., & Russell, J. A.: Judgments of emotion from spontaneous facial expressions of New Guineans. Emotion 7: 736–744. (2007)

Nass, C., Foehr, U. and Somoza, M.: The effects of emotion of voice in synthesized and recorded speech, (Eds.), Proceedings of the AAAI Symposium Emotional and Intelligent II: The Tangled Knot of Social Cognition, North Falmouth, MA. (2001)

Niedenthal, P. M., Halberstadt, J. B., Margolin, J., & Innes-Ker, A. H.: Emotional state and the detection of change in facial expression of emotion. Eur. J. Soc. Psychol. 30:211–22. (2000)

Niedenthal, P. M., Barsalou, L. W., Ric, F., & Krauth-Gruber, S.: in Emotion: Conscious and Unconscious, L. F. Barrett, P. M. Niedenthal, P. Winkielman, Eds. (Guilford, New York, 2005), pp. 21–50. (2005)

Niedenthal, P. M., Kruth-Gruber, S., & Ric, F.: Theories of Emotion. The Psychology of Emotion: Interpersonal Experiential, and Cognitive Approaches, Principles of Social Psychology series, New York: Psychology Press, pp.136-144. (2006a)

Niedenthal, P. M., Kruth-Gruber, S., & Ric, F.: What Information Determines the Recognition of Emotion? The Psychology of Emotion: Interpersonal Experiential, and Cognitive Approaches, Principles of Social Psychology series, New York: Psychology Press, pp.136-144. (2006b)

Niedenthal, et al, Embodying emotion, Science, VOL 316, 2007. (2007)

Niedenthal, P. M., Mondillon, L., Winkielman, P., & Vermeulen, N.: Embodiment of emotion concepts, J Pers Soc Psychol, 2009;96:1120–1136. (2009)

Noël, S., Dumoulin, S., & Lindgaard, G.: Interpreting Human and Avatar Facial Expressions, in Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R., and Winckler, M. (Eds.): Human-Computer Interaction – INTERACT 2009 (Springer Berlin /Heidelberg, 2009), pp. 98-110. (2009)

Oatley, K., & Jenkins, J. M.: Understanding emotions. Oxford: Blackwell. (1996)

Ogura, Y., Aikawa, H., Shimomura, K., Kondo, H., Morishima, A., Lim, H., & Takanishi, A.: Development of a new humanoid robot WABIAN-2, Proc. IEEE Int. Conf. Robot, Autom, pp. 76–81. (2006)

Ogata, T., Okuno, H. G., & Kozima, H.: Motion from sound: Intermodal neural network mapping, IEEE Intelligent Systems, Vol.23, No.2, pp.76-78. (2008)

Pantic, M. & Rothkrantz, L. J. M.: Toward an affect-sensitive multimodal human-computer interaction, Proceedings of the IEEE, Volume 91, Issue 9, Page(s): 1370 – 1390, Sept. (2003)

Parkinson, B.: Untangling the appraisal-emotion connection. Personality and Social Psychology Review, 1:62–79. (1997)

Patrick Rau, P. L., Li, Y., & Li, D.: Effects of communication style and culture on ability to accept recommendations from robots, Computers in Human Behavior 25, 587–595. (2009)

Pecher, D., Zeelenberg, R., & Barsalou, L. W.: Verifying different-modality properties for concepts produces switching costs, Psychol Sci, 14(2):119-24, 2003 Mar. (2003)

Picard, R.W.: Affective Computing. M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 321. (1995)

Picard, R.W.: Affective Computing for HCI. Proceedings HCI, Munich, Germany. (1999)

Picard, R.W.: Affective computing: challenges, Int. J. Human-Computer Studies 59, 55–64. (2003)

Plutchik, R.: Emotion: A psychoevolutionary analysis. New York, NY: Harper & Row. (1980)

Plutchik, R.: Emotions: A general psychoevolutionary theory. In K. R. Scherer & P. Ekman (Eds.), Approaches to emotion (pp. 197–219). Hillsdale, NJ: Lawrence Erlbaum. (1984)

Pollak, S. D., & Kistler, D. J.: Early experience is associated with the development of categorical representations for facial expressions of emotion. In Proceedings of the national academy of sciences of the united states of America, 99(13), 9072–9076. (2002)

Pollak, S. D., Messner, M., Kistler, D.J., & Cohn, J.F.: Development of perceptual expertise in emotion recognition. Cognition, 110, 242–247. (2009)

Pollak, S. D., & Sinha, P.: Effects of early experience on children's recognition of facial displays of emotion. Developmental Psychology, 38, 784–791. (2002)

Posner, J., Russell, J., & Peterson, B.: The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev Psychopathol 17(03):715–734. (2005)

Powers, W. T., Clark, R. K., & McFarland, R. L.: A general feedback theory of human behavior: Part I.Perceptual and Motor Skills, 11, 71-88. (1960)

Powers, W. T.: Behavior: The control of perception (2nd ed.). Chicago: Aldine de Gruyter. (1973)

Powers, W. T.: On Emotions and PCT: A brief overview. In Perceptual Control Theory: Science & Applications - A Book of Readings. Living Control Systems Publishing. (2007)

Powers, W. T.: Living Control Systems III: The Fact of Control. Benchmark Publications. (2008)

Prendinger, H., Mayer, S., Mori, J., & Ishizuka, M.: Persona effect revisited. Using bio-signals to measure and reflect the impact of character-based interfaces. In: Rist,T., Aylett, R., Ballin, D., Rickel, J. (Eds.), Fourth International Working Conference On Intelligent Virtual Agents (IVA03). Springer, KlosterIrsee, Germany, pp.283–291. (2003)

Prinz, J. J.: Gut reactions: A perceptual theory of emotion. New York, NY: Oxford University Press. (2004a)

Prinz, J. J.: Which Emotions Are Basic? In D. Evans and P. Cruse (Eds.), Emotion, Evolution, and Rationality, Oxford University Press. (2004b)

Reisenzein, R.: Arnold's theory of emotion in historical perspective. Cognition and Emotion, 20, 920-951. (2006)

Reeves, B., & Nass, C.: The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge: Cambridge University Press. (1996)

Roberts, N. A., Tsai, J. L. & Coan, J. A.: Emotion elicitation using dyadic interaction tasks. In J. A. Coan and J. J. B. Allen (Eds.) Handbook of Emotion Elicitation and Assessment (pp. 106-123), New York, NY: Oxford University Press. (2007)

Roccella, S., Carrozza, M. C., Cappiello, G., Dario, P., Cabibihan, J., Zecca, M., Miwa, H., Itoh, K., Matsumoto, M., & Takanishi, A.: Design, fabrication and preliminary results of a novel anthropomorphic hand for humanoid robotics: RCH-1, Proc. 2004IEEE/RSJ Int. Conf. Intell. Robots Syst. (Sendai) pp. 266–271. (2004)

Rolls, E.T.: A theory of emotion, and its application to understanding the neuralbasis of emotion. In Oomura,Y. (ed.) Emotions. Neural and Chemical Control, pp. 325–344.Basel: Karger. (1986a)

Rolls, E.T.: Neural systems involved in emotion in primates. In Plutchik, R. andKellerman, H. (eds) Emotion: Theory, Research, and Experience. Vol. 3. BiologicalFoundations of Emotion, pp. 125–143. New York: Academic Press. (1986b)

Rolls, E.T.: A theory of emotion, and its application to understanding the neuralbasis of emotion. Cognition and Emotion 4, 161–190. (1990)

Rolls, E.T.: The Brain and Emotion. Oxford: Oxford University Press. (1999)

Rolls, E.T.: Précis of The Brain and Emotion. Behavioral and Brain Sciences 23,177–233. (2000)

Rolls, E.T.: Emotion Explained. Oxford: Oxford University Press. (2005)

Rolls, E. T.: A neurobiological approach to emotional intelligence. Chapter 3, pp. 72-100 in The Science of Emotional Intelligence, Eds. G.Matthews, M.Zeidner and R.D.Roberts. Oxford University Press: Oxford. (2007)

Rolls, E. T., & Stringer, S. M.: A model of the interaction between mood and memory.Network: Computation in Neural Systems, 12, 111–129. (2001)

Rolls, E. T., & Treves, A.: Neural networks and brain function. Oxford, UK: Oxford University Press. (1998)

Roseman, I. J. & Smith, C. A.: Appraisal theory: overview, assumptions, varieties, controversies. In Appraisal processes in emotion (eds K. R. Scherer, A. Schorr &T. Johnstone), pp. 3–34. New York, NY: Oxford University Press. (2001)

Rottenberg, J., Ray, R. R., & Gross, J. J.: Emotion elicitation using films. In J.A. Coan & J.J.B Allen (Eds.), Handbook of Emotion Elicitation and Assessment (pp. 9-28). New York: Oxford University Press. (2007)

Rozin, P., Lowery, L., Imada, S., & Haidt, J.: The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). Journal of Personality and Social Psychology, 76, 574-586. (1999)

Ruef, A. M., & Levenson, R. W.: Continuous measurement of emotion-the affect rating dial. In J. A. Coan and J. J. B. Allen (Eds.) Handbook of Emotion Elicitation and Assessment (pp. 286-297), New York, NY: Oxford University Press. (2007)

Russell, J. A.: Is there universal recognition of emotion from facial expressions? A review of cross-cultural studies. Psychol. Bull. 115:102–41. (1994)

Russell, J.: Reading emotions from and into faces: resurrecting a dimensional–contextual perspective. In: Russell, J., Fernandez-Dols, J. (Eds.), The Psychology of Facial Expression. Cambridge University Press, Cambridge, UK, pp. 295–320. (1997)

Russell, J. A.: Core affect and the psychological construction of emotion. Psychol. Rev. 110, 145–172. (2003)

Russell, J. A.: Emotion, core affect, and psychological construction. Cognition & Emotion, 23:7(1):1259–1283. (2009)

Russell, J. A., Bachorowski, Jo-Anne, & Fernández-Dols, José-Miguel: Facial and Vocal Expressions of Emotion. Annu. Rev. Psychol. 2003. 54:329–49. (2003)

Russell, J. A, & Fernández-Dols, J. M.: What does a facial expression mean? See Russell & Fernández-Dols 1997a, pp. 3–30. (1997b)

Russell, J. A., & Fehr, B.: Relativity in the perception of facial expression. In Journal of Experimental Psychology: General, 115, 102-141. (1987)

Russell, J. A., Weiss, A., Mendelsohn, G. A.: Affect Grid: A Single-Item Scale of Pleasure and Arousal. Journal of Personality and Social Psychology, Vol. 57, No. 3, pp. 493-502 (1989)

Sander, D., Grandjean, D. & Scherer, K. R.: A systems approach to appraisal mechanisms in emotion. Neural Networks 18, 317–352. (2005)

Saint-Aimé, S., Le-Pévédic, B., & Duhaut, D.: iGrace – Emotional Computational Model for EmI Companion Robot. In Advances in Human-Robot Interaction, edited by Vladimir A. Kulyukin, INTECH, Croatia. (2009)

Saldien, J., Goris K., Vanderborght B., Vanderfaeillie J., & Lefeber D.: Expressing emotions with the social robot Probo. Int J Soc Robot (2010) 2: 377–389. (2010)

Saldien, J., Goris, K., Vanderborght, B., Verrelst, B., Van Ham, R., & Lefeber, D.: ANTY: The development of an intelligent huggable robotfor hospitalized children. CLAWAR. (2006)

Saldien, J., Goris, K., Yilmazyildiz, S., Verhelst, W., Lefeber, D.: On the design of the huggable robot Probo. Journal of Physical Agents, Special Issue on Human Interaction with Domestic Robots, Vol 2, No 2, pp. 3-11. (2008)

Scarantino, A.: Core Affect and Natural Affective Kinds. Philosophy of Science, 76 (December 2009), pp. 940–957. (2009)

Schachter, S. & Singer, J. E.: Cognitive, Social, and Physiological Determinants of Emotional State. Psychological Review, 69(5), 379-399. (1962)

Scherer, K. R.: On the nature and function of emotion: a component process approach. In Approaches to emotion (eds K. R. Scherer & P. Ekman), pp. 293–317. Hillsdale, NJ: Erlbaum. (1984a)

Scherer, K. R.: Emotion as a multi-component process: A model and some cross-cultural data. In P. Shaver, Review of personality and social psychology. Emotions, relationships and health (Vol. 5) (pp. 37–63). Beverly Hills, CA: Sage, 37–63. (1984b)

Scherer, K. R.: Vocal affect expression: A review and a model for future research. Psychological Bulletin, 99, 143–165. (1986)

Scherer, K. R.: Studying the emotion-antecedent appraisal process: An expert system approach. Cognition and Emotion, 7, 325–355. (1993)

Scherer, K. R.: Profiles of emotion-antecedent appraisal: Testing theoretical predictions across cultures. Cognition and Emotion, 11,113–150. (1997)

Scherer, K. R.: Appraisal theories. In Handbook of cognition and emotion (eds T. Dalgleish & M. Power), pp. 637–663. Chichester, UK: Wiley. (1999)

Scherer, K. R.: Appraisal considered as a process of multilevel sequential checking. In Appraisal processes in emotion: theory, methods, research (eds K. R. Scherer, A.Schorr & T. Johnstone), pp. 92–120. New York, NY: Oxford University Press. (2001)

Scherer, K. R.: What are emotions? And how can they be measured?. Social Science Information, 44(4), 695-729. (2005)

Scherer, K. R., Banse, R., & Wallbott, H. G.: Emotion inferences from vocal expression correlate across languages and cultures. J. Cross-Cult. Psychol. 32:76–92. (2001)

Scherer, K. R.: Emotions are emergent processes: They require a dynamic computational architecture. Philosophical Transactions of the Royal Society, Series B, 364, 3459-3474. (2009)

Schiffbauer, A.: Effect of Observer's emotional state on judgments of the emotional state of others. Journal of Personality and Social Psychology, 30, 31-35. (1974)

Shimada, M., Minato, T., Itakura, S., Ishiguro, H.: Evaluation of Android Using Unconscious Recognition, Proc. 6th IEEE-RAS Int. Conf. Humano. Robots. pp. 157–162. (2006)

Shin, L. M., Dougherty, D., Macklin, M. L., Orr, S. P., Pitman, R. K., & Rauch, S.L.: Activation of anterior paralimbic structures during guilt-related script-driven imagery. Biological Psychiatry, 48, 43-50.

(2000)

Sidner, C.L., Lee, C., Kidd, C. D., Lesh, N., & Rich, C.: Explorations in engagement for humans and robots, Artif. Intell. 166(1-2), 140–164. (2005)

Smith, C. A. & Ellsworth, P. C.: Patterns of cognitive appraisal in emotion. J. Pers. Soc. Psychol. 48, 813–838. (1985)

Smith, C., & Scott, H: A componential approach to the meaning of facial expressions, in J. Russell and J. Fernandez-Dols, eds., "The Psychology of Facial Expression," Cambridge University Press, Cambridge, UK, pp. 229–254. (1997)

Solis, J., Chida, K., Suefuji, K., & Takanishi, A.: The development of the anthropomorphic flutist robot at Waseda University, Int. J. Humano. Robots 3(2), 1–25. (2006)

Sosnowski, S., Bittermann, A., Kuhnlenz, K., & Buss, M.: Design and evaluation of emotion-display EDDIE. In: IEEE/RSJ international conference on intelligent robots and systems, pp 3113–3118. (2006)

Spence, C., Nicholls M. E., & Driver, J.: The cost of expecting events in the wrong modality, Percept. Psychophys, 2001, 63:330–336. (2001)

Stein, N. L., & Hernandez, M. W.: Assessing understanding and appraisals during emotional experience: The development and use of the Narcoder. In J. A. Coan and J. J. B. Allen (Eds.) Handbook of Emotion Elicitation and Assessment (pp. 298-317), New York, NY: Oxford University Press. (2007)

Stein, N.L., & Trabasso, T.: The organization of emotional experience: creating links among emotion, thinking, language, and intentional action. Cognition & Emotion, 6(3-4), 225-244. (1992)

Tanaka, F., Noda, K., Sawada, T., Fujita M.: Associated emotion and its expression in an entertainment robot QRIO, Proc. 3rd Int. Conf. Entertain. Comput. (Eindhoven) pp. 499–504. (2004)

Tao, J., Tan, T., & Picard (Eds.), R.W.: Affective Computing: A Review, ACII 2005, LNCS3784, pp. 981 – 995. (2005)

Thayer, S: The effect of facial expression sequence upon judgments of emotion. Journal of Social Psychology, 111, 305-306. (1980a)

Thayer, S: The effect of expression sequence and expressor identity on judgments of the intensity of facial expression. Journal of Nonverbal Behavior, Vol 5(2), 71-79. (1980b)

Thórisson, K. R.: Natural turn-taking needs no manual: Computational theory and model, from perception to action, Multimodality in Language and Speech Systems, pp. 173-207. (2002)

Carey, T.: PCT: A right turn for researchers — PCT50. In http://www.pctweb.org/carey.pdf

Tomkins, S. S.: Affect. Imagery. Consciousness (Vol. 1). The positive affects. New York: Springer Publishing Co. (1962)

Tomkins, S. S.: Affect. Imagery. Consciousness (Vol. 2). The negative affects. New York: Springer Publishing Co. (1963)

Tooby, J., & Cosmides, L.: The past explains the present: Emotional adaptations and the structure of ancestral environments. Ethology and Sociobiology, 11, 375-424. (1990)

Ursula, H., Reginald, B. A., Jr, & Robert, E. K.: The face is not an empty canvas: how facial expressions interact with facial appearance. Phil. Trans. R. Soc. B December 12, 2009, 364:3497-3504. (2009)

Van Breemen, A., Yan, X., & Meerbeek, B.: iCat: an animated user-interface robot with personality. In Autonomous Agents and Multi agent Systems, AAMAS'05.ACM Press, New York, NY, USA, 143-144. (2005)

Vasquez, K., Keltner, D., Ebenbach, D. H., & Banaszynski, T. L.: Cultural variation and similarity in moral rhetorics: Voices from the Philippines and United States. Journal of Cross-Cultural Psychology, 32, 93-120. (2001)

Vermeulen, N., Niedenthal, P. M., & Luminet, O.: Switching between sensory and affective systems incurs processing costs, Cogn Sci, 31:183–192, 2007. (2003)

Wada, K., Shibata, T., Sakamoto, K., & Tanie, K.: Long term interaction between seal robots and elderly people – robot assisted activity at a health service facility for the aged, Proc. 3rd Int. Symp. Auton. Minirobots Res. Edutain. pp. 325–330. (2005)

Wahlster, W.: Towards Symmetric Multimodality: Fusion and Fission of Speech, Gesture, and Facial Expression, LECTURE NOTES IN COMPUTER SCIENCE, Springer. (2003)

Wallbott, H. G.: Faces in context: the relative importance of facial expressions and context in formation in determining emotion attributions. In Facets of Emotion: Recent Research (K.R.Scherer, ed.), pp.139-161. Erlbaum; Hillsdale. (1988)

Walters, M. L.: The design space for robot appearance and behaviour for social robot companions, Doctoral Thesis, University of Hertfordshire. (2008)

Walters, M. L., Syrdal, D. S., Dautenhahn, K., Boekhorst, R. te, & Koay, K. L.: Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion, Auton Robots, 24(2):159–178. (2008)

Watson, D., Clark, L. A., & Tellegen, A.: Development and validation of brief measures of positive and negative affect: The PANAS Scales. Journal of Personality and Social Psychology, 47, 1063–1070. (1988)

Watson, S. G.: Judgment of emotion from facial and contextual cue combinations. Journal of Personality and Social Psychology, 24, 334-342. (1972)

Weiskrantz, L.: Emotion. In Weiskrantz, L. (ed.) Analysis of Behavioural Change,pp. 50–90. New York: Harper and Row. (1968)

Widen, S. C., & Russell, J. A.: Gender and preschoolers perception of emotion. Merrill- Palmer Q. 48:248–62. (2002)

Russell, J. A.: Culture and the categorization of emotion. Psychol. Bull. 110:426– 50. (1991)

Yamaoka, F., Kanda, T., Ishiguro, H., & Hagita, N.: How contingent should a lifelike robot be? The relationship between contingency and complexity. Connection Science, 19, 2, 143-162. (2007)

Yuki, M., Maddux, W., & Masuda, T.: Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States. Journal of Experimental Social Psychology, 303-311. (2007)

Appendix Ⅰ: A List of the Actual Order of the 11 Experiments

Table Appendix Ⅰ The actual order of the 11 experiments

Experimental Order	Context Types	Manner Of Presentation (simultaneous/separate)	Viewing Order (first/second)
Speech1	News Recordings	Simultaneous	First
Image1	Affective Pictures	Simultaneous	First
Music3	Musical Pieces (20mins)	Separate	Second
Image3	Affective Pictures	Separate	First
Video2	Film Clips	Simultaneous	Second
Video1	Film Clips	Simultaneous	First
Music2	Musical Pieces (3mins)	Simultaneous	Second
Music1	Musical Pieces (3mins)	Simultaneous	First
Image2	Affective Pictures	Simultaneous	Second
Video3	Film Clips	Separate	First
Speech2	News Recordings	Simultaneous	Second

Appendix Ⅱ: Detailed Experimental Set Ups for All the 11 Experiments

Table Appendix Ⅱ More experimental set ups for all the 11 experiments

<table>
<tr><th>Experiment (date)</th><th>Date Venue and Room Condition</th><th>Experimental Facilities</th></tr>
<tr><td>Speech1 (31/03/2010 -14/05/2010)</td><td>Room G35 (a large lab contains four other researchers), Regents Court, the University of Sheffield. The room is warm, quiet and bright, although it could be noisy if other researchers talked loudly.</td><td>Each participant was seated on a high back office chair, in front of the left-hand side of a long rectangular desk (the experimenter was seated in front of the right-hand side of the desk). Each participant could adjust the height of the chair and distance between the chair and the desk such that he/she could view the robot head on the left-hand side of the desk properly. Each participant could also see a computer screen next to the left-hand side of the robot head. During each experimental section, each participant could listen to a piece of recorded BBC News played by the computer, through a stereo headphone, but he/she could not operate the computer. By selecting the options on the control panel written in Java on the computer screen, the experimenter decided which robotic facial expressions each participant would watch and which piece of BBC News the participant would listen to. Although the participant witnessed the selection process, no names of the facial expressions or the News were revealed.</td></tr>
<tr><td>Image1 and Music3 (14/10/2010-12/11/2010)</td><td rowspan="5">Room G105 (a staff office), Regents Court, the University of Sheffield. The room is warm and bright, and most importantly, it is very quiet as during each experimental section, there was only a participant and the experimental conductor inside the room.</td><td rowspan="5">Compared with that in the first experiment, each experimental section took place in a smaller room, with a different rectangular desk and a different high back office chair (again, adjustable), but with the same robot head (on the left-hand side of the desk), the same computer screen (on the right-hand side of the desk, next to the robot head), and the same headphones. Also differently, the experimenter was seated in front of another desk, the vertical line of which closely parallels to the horizontal line of the desk placed by the robot head. Each participant could listen to the News recording or the musical piece played by the computer through the headphone, and could watch the affective pictures or film clip on the computer screen. However, for most of the time of each section, the participant could only see the back of the computer screen, unless the experimental conductor turned the screen to face the participant to enable the participant to see some affective pictures or a film clip. Again, the participant could not operate the computer, and no names of the facial expressions or the News, or the music, or the affective pictures, or the film clips would be revealed.</td></tr>
<tr><td>Image3 and Video2 (16/02/2011-11/03/2011)</td></tr>
<tr><td>Video1 and Music2 (17/03/2011-25/03/2011)</td></tr>
<tr><td>Music1 and Image2 (25/03/2011-30/03/2011)</td></tr>
<tr><td>Video3 and Speech2 (30/03/2011-08/04/2011)</td></tr>
</table>

Appendix Ⅲ: The Questionnaire of the Experiment Speech1

An Investigation of How Users Perceive the Facial Expressions of an Emotional Robot

Jiaming Zhang

Dr. Amanda Sharkey, Prof. Noel Sharkey

Neurocomputing and Robotics Group

Department of Computer Science, the University of Sheffield, UK

Information Sheet

1. Research Project Title:

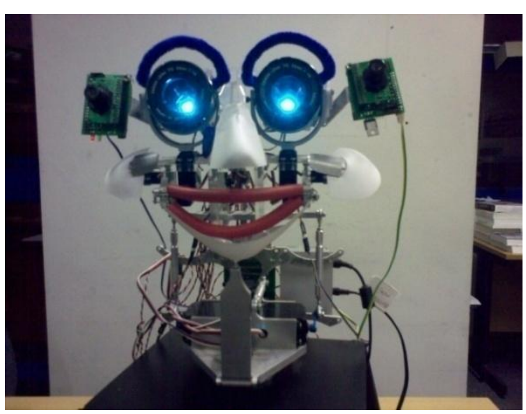
An Investigation of How Users Perceive the Facial Expressions of an Emotional Robot

2. Invitation paragraph

You are being invited to take part in a research project. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

3. What is the project's purpose?

In this project, a mechanical robot head called CIM (see the picture on the next page) in the Neurocomputing and Robotics lab in the University of Sheffield, was programmed to express various simulated emotions, such as Paul Ekman' six universal emotions: happiness, sadness, anger, disgust, surprise, and fear. The main purpose of this project is to find out more about how people perceive the facial expressions of this emotional robot. We are interested in whether the context of a facial expression affects how it is seen, and whether the facial expressions of a robot can affect people's opinions about robots.



4. What is the overview of the project procedures?

Firstly, an emotional robot will show six different facial expressions to all subjects for warming up purpose and subjects will fill in the first part of the questionnaire of how people perceive a robot's simulated emotions. Secondly, subjects will listen to a piece of recorded BBC News while being shown a corresponding sequence of facial expressions of the emotional robot. After listening to the recorded BBC News, subjects will fill in the second part of the questionnaire that cover two aspects: their views of the emotions that the robot was expressing during the recorded news, their moods, respectively. Further explanation of the project can be given if required at the end.

5. Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep (and be asked to sign a consent form) and you can still withdraw at any time without it affecting any benefits that you are entitled to in any way. You do not have to give a reason.

6. Will my taking part in this project be kept confidential?

All the information that we collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified in any reports or publications.

Thank you very much for taking part in this project!

Date: 1st, April, 2010

Name of Applicant: Jiaming Zhang

Participant Consent Form

Title of Research Project: An Investigation of How Users Perceive the Facial Expressions of an Emotional Robot

Name of Researcher: Jiaming Zhang

Participant Identification Number for this project: **Please initial box**

1. I confirm that I have read and understand the information sheet dated *[* 1st, April, 2010*]* explaining the above research project and I have had the opportunity to ask questions about the project. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason and without there being any negative consequences. In addition, should I not wish to answer any particular question or questions, I am free to decline. ☐

3. I understand that my responses will be kept strictly confidential. I give permission for members of the research team to have access to my anonymised responses. I understand that my name will not be linked with the research materials, and I will not be identified or identifiable in the report or reports that result from the research. ☐

4. I agree for the data collected from me to be used in future research ☐

5. I agree to take part in the above research project. ☐

________________	________________	________________
Name of Participant	Date	Signature
________________	________________	________________
Lead Researcher	Date	Signature

Participant's Information

Male/Female:__________

Age:________

Nationality:_____________

Educational Level:____________

Research Topic (if appropriate):___________________________________

How interested are you in robots? _______

A: Very interested B: Interested

C: Neither interested nor not interested

D: Not interested E: Not at all interested

How good are you at recognising people's emotions? _______

A: Very good B: Good C: Average

D: Poor E: Very Poor

Project Procedures

1. The beginning of this project is only for the purpose of warming up. You will be shown six different facial expressions of an emotional robot. After each facial expression is shown, please select what emotion you think the robot was feeling from the given choices (Page 7). This procedure will take less than 5 minutes.
2. Please listen to a piece of recorded BBC News. When you are listening, please pay attention to the robot head on the desk, as it will show you a corresponding sequence of facial expressions. This procedure will take less than 5 minutes.
3. Please fill in Aspect One (Page 8).
4. Please fill in Aspect Two (Page 9).

Now, please look at the robot head in front of you, and turn to page 7.

Warm Up

1. Please select what emotion you think the robot was feeling from the following given choices. ____________

A: Joy B: Sadness C: Anger D: Fear E: Disgust F: Surprise

2. Please select what emotion you think the robot was feeling from the following given choices.____________

A: Joy B: Sadness C: Anger D: Fear E: Disgust F: Surprise

3. Please select what emotion you think the robot was feeling from the following given choices. ____________

A: Joy B: Sadness C: Anger D: Fear E: Disgust F: Surprise

4. Please select what emotion you think the robot was feeling from the following given choices.____________

A: Joy B: Sadness C: Anger D: Fear E: Disgust F: Surprise

5. Please select what emotion you think the robot was feeling from the following given choices. ____________

A: Joy B: Sadness C: Anger D: Fear E: Disgust F: Surprise

6. Please select what emotion you think the robot was feeling from the following given choices. ____________

A: Joy B: Sadness C: Anger D: Fear E: Disgust F: Surprise

Now, please put on a headphone and listen to a piece of recorded BBC News, and then fill in the following four aspects of the rest of the questionnaire.

Aspect One

1. As a total impression, please select what kind of News you think you were listening to from the following given choices. ___________

A: Positive News   B: Neutral News   C: Negative News

2. As a total impression, please select what kind of emotion (affect) you think the robot was feeling from the following given choices.___________

A: Positive Affect   B: Neutral Affect   C: Negative Affect

3. As a detailed impression, please select what emotions you think the robot were feeling from the following given choices. ___________(Multiple choices)

A: Joy   B: Sadness   C: Anger   D: Fear   E: Disgust   F: Surprise

Aspect Two

INSTRUCTIONS: Circle the response on the scale below that indicates how well each adjective or phrase describes your present mood.

(definitely do not feel) (do not feel) (slightly feel) (definitely feel)

XX X V VV

Lively	XX X V VV	Drowsy	XX X V VV
Happy	XX X V VV	Grouchy	XX X V VV
Sad	XX X V VV	Peppy	XX X V VV
Tired	XX X V VV	Nervous	XX X V VV
Caring	XX X V VV	Calm	XX X V VV
Content	XX X V VV	Loving	XX X V VV
Gloomy	XX X V VV	Fed up	XX X V VV
Jittery	XX X V VV	Active	XX X V VV

Appendix Ⅳ: The Procedures of Experiments Image1 and Music3 Conducted in Pairs

Experiments Image1 and Music3 were conducted in pairs. In Image1, a series of affective pictures was presented at the same time as the sequence of robot expressions. In Music3, a piece of classical music was presented before the sequence of synthetic robot emotional expressions. The two sequences of facial expressions (positive and negative) of the emotional robot used in the two experiments were the same as those used in experiment Speech1. The classical music used in Music3 was about 20 minutes long.

Subjects in the beginning of experiment Image1 read similar information sheet as instructions and filled in a similar questionnaire as provided in the experiment Speech1 (see Appendix Ⅲ: The Questionnaire of the Experiment Speech1). For instance, the overview of the project procedures provided in the information sheet of the second experiment was presented as follows: *Firstly, you will be asked to read some Instructions of Affect Grid which will be used later in the study. And, to warm up, you will be shown a few different facial expressions of an emotional robot and slides of affective pictures. Secondly, you will be asked to watch slides of affective pictures on a computer screen while being shown a corresponding sequence of facial expressions of the emotional robot, and then to fill in a short questionnaire. In the third part of the study, you will be asked to relax and listen to some classical music. As you listen to the music, you will be asked to fill in Affect Grids. After listening to the music, you will be asked to watch the robot head and to fill in a final short questionnaire. There will only be time for a short explanation of the project at the end, further details or discussion will have to take place via emails.*

Then, he/she went through the following procedures:

Warm up: Six different facial expressions of the emotional robot CIM and six static affective pictures on a computer screen selected from the IAPS were showed simultaneously to all subjects. The six pictures are all with neutral valence, i.e., they are a spoon, a blue cup, a cliff hanger drinking a beer while hanging on a cliff, a boy thinking next to a chess board, a mushroom, and an erupting volcano, respectively. Then, as in the first experiment, subjects were asked to fill in a questionnaire about how they perceived the robot's static facial expressions. This time, before the warming-up procedure, subjects were also asked to read the instructions (see the Appendix in (Russell et al. 1989)) for the Affect Grid used to measure current levels of pleasure and arousal, as it would be used later in the experiment.

Experiment Image1: Subjects watched 32 slides of affective pictures while being simultaneously shown a 3-minute sequence of facial expressions (either positive or negative) of the emotional robot.

Responses for Image1: After the last picture was shown, subjects were asked to use the Affect Grid to indicate their current state, before they answered the following questions:

1. As a total impression, please select what kind of affective pictures you think you were viewing from the following given choices. ___________

A: Pleasant Picture B: Neutral Pictures C: Unpleasant Pictures

2. As a total impression, please select what kind of emotion (affect) you think the robot was feeling from the following given choices.___________

A: Positive Affect B: Neutral Affect C: Negative Affect

3. As a detailed impression, please select what emotions you think the robot was feeling from the following given choices. ___(you can choose more than one option)

A: Joy B: Sadness C: Anger D: Caring E: Disgust F: Surprise

Experiment Music3: After completing experiment Image1 subjects were played a piece of classical music that was happy or sad (both were about 20 minutes). Without an instruction, subjects were asked by the experimental conductor, to mark the Affect Grid twice while listening to the music (after 10 minutes and after 20 minutes of music had played). After the music, subjects were showed a 3-minute sequence of facial expressions of the emotional robot with the opposite valance to that used in experiment Image1.

Responses for Music3: After being shown the sequence of the synthetic robot emotions, subjects were asked to answer the following questions:

1. As a total impression, please select what kind of emotional music you think you were listening to from the following given choices. ___________

A: Happy Music B: Neutral Music C: Sad Music

2. As a total impression, please select what kind of emotion (affect) you think the robot was feeling from the following given choices.___________

A: Positive Affect B: Neutral Affect C: Negative Affect

3. As a detailed impression, please select what emotions you think the robot was feeling from the following given choices. ___(you can choose more than one option)

A: Joy B: Sadness C: Anger D: Caring E: Disgust F: Surprise

Appendix Ⅴ: The Procedures of Experiments Image3 and Video2 Conducted in Pairs

Experiments Image3 and Video2 were conducted in pairs. In Image3, a series of affective pictures was presented before the sequence of synthetic robot emotional expressions. In Video2, a film clip was presented at the same time as the sequence of robot expressions. The two sequences of facial expressions (positive and negative) of the emotional robot used in the two experiments were the same as those used in experiment Speech1.

Subjects in the beginning of experiment Image3 read similar information sheet as instructions and filled in a similar questionnaire as provided in the first experiment (see Appendix Ⅲ: The Questionnaire of the Experiment Speech1). For instance, the overview of the project procedures provided in the information sheet of the third experiment was presented as follows:

Firstly, you will be asked to read some Instructions of Affect Grid which will be used later in the study. And, to warm up, you will be shown a few different facial expressions of an emotional robot. Secondly, you will be asked to watch slides of affective pictures on a computer screen before being shown a corresponding sequence of facial expressions of the emotional robot, and then to fill in a short questionnaire. In the third part of the study, you will be asked to watch a film clip while being shown a corresponding sequence of facial expressions of the same robot head and to fill in a final short questionnaire. There will be no time for an explanation of the project at the end, further details or discussion will have to take place via emails.

Then, he/she went through the following procedures:

Warm up: Six different facial expressions of the emotional robot CIM were showed all subjects, after subjects were asked to read the instructions for the Affect Grid which would be used later in the experiment.

Experiment Image3: Subjects watched 32 slides of affective pictures before being shown a 3-minute sequence of facial expressions (either positive or negative) of the emotional robot.

Responses for Image3: After the robot emotions, subjects were asked to use the Affect Grid to indicate their current state, before they answered the following questions:

1. As a total impression, please select what kind of affective pictures you think you were viewing from the following given choices. ___________

A: Pleasant Picture   B: Neutral Pictures   C: Unpleasant Pictures

2. As a total impression, please select what kind of emotion (affect) you think the robot was feeling from the following given choices.___________

A: Positive Affect   B: Neutral Affect   C: Negative Affect

3. As a detailed impression, please select what emotions you think the robot was feeling from the following given choices. ___(you can choose more than one option)

A: Joy B: Sadness C: Anger D: Caring E: Disgust F: Surprise

Experiment Video2: After filling in the last question in Image3, subjects watched a film clip and a 3-minute sequence of facial expressions of the emotional robot with the opposite valance in Image3 simultaneously.

Responses for Video2: After the synthetic robot emotions, subjects were asked to fill in a copy of Affect Grid before answering the following questions:

1. As a total impression, please select what kind of emotional film clip you think you was watching from the following given choices. ___________

A: Amused (Happy) Film Clip B: Neutral Film Clip C: Sad Film Clip

2. As a total impression, please select what kind of emotion (affect) you think the robot was feeling from the following given choices.___________

A: Positive Affect B: Neutral Affect C: Negative Affect

3. As a detailed impression, please select what emotions you think the robot was feeling from the following given choices. ___(you can choose more than one option)

A: Joy B: Sadness C: Anger D: Caring E: Disgust F: Surprise

Appendix Ⅵ: The Procedures of Experiments Video1 and Music2 Conducted in Pairs

Experiments Video1 and Music2 were conducted in pairs. In Video1, a film clip was presented at the same time as the sequence of robot expressions. In Music2, a piece of classical music was presented with the robot behaviour simultaneously. The two sequences of facial expressions (positive and negative) of the emotional robot used in the two experiments were the same as those used in experiment Speech1. The film clips were as the same as that in experiment Video2. The classical music used this time was only 3 minutes long.

Subjects in the beginning of experiment Video1 filled in a similar questionnaire as provided in previous experiments, without an information sheet as instructions. Rather, to save time and fit more participants into the slots of the experimental schedule, they read the overview of the project procedures provided in a piece of paper (see Fig. Appendix Ⅵ, in both English and Chinese, as there were lots of Chinese participants), before the warm up procedure as follows:

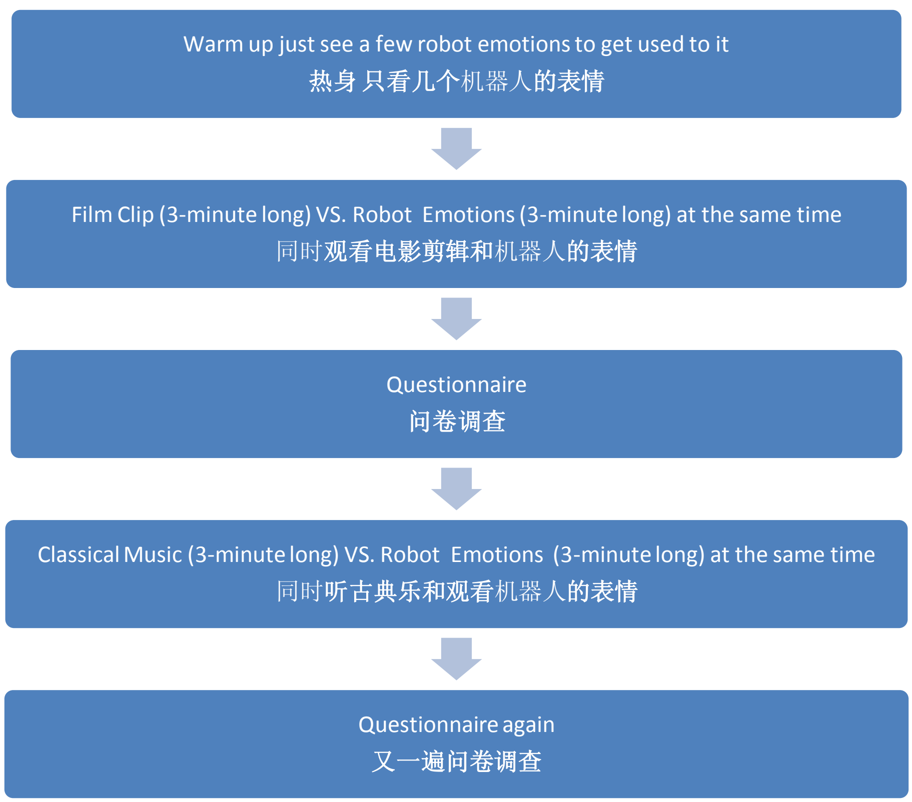


Fig. Appendix Ⅵ: Experimental Instructions of the two experiments Video1 and Music2

When the participant was reading the above paper, he/she received a verbal description in English (a Chinese version can be provided if a Chinese participant requested) of the whole

procedures from the experimental conductor, e.g., as follows:

Firstly, you will watch six different facial expressions of an emotional robot to warm you up. Secondly, you will watch a film clip and a corresponding sequence of facial expressions of the robot head simultaneously. Then you will need to fill in a short questionnaire. In the third part of the study, you will listen to a piece of classical music and watch a corresponding sequence of facial expressions of the same robot at the same time. Finally, you will need to fill in a final questionnaire. But before the experiment, you need to read some Instructions of Affect Grid which will be used later in the study.

Then, he/she went through the following procedures:

Warm up: Six different facial expressions of the emotional robot CIM were shown to all subjects, after subjects were asked to read the instructions for the Affect Grid which would be used later in the experiment.

Experiment Video1: Subjects watched a film clip while being shown a 3-minute sequence of facial expressions (either positive or negative) of the emotional robot simultaneously.

Responses for Video1: After the robot emotions, subjects were asked to use the Affect Grid to indicate their current state, before they answered the following questions:

1. As a total impression, please select what kind of emotional film clip you think you were watching from the following given choices. ___________

A: Amused (Happy) Film Clip B: Neutral Film Clip C: Sad Film Clip

2. As a total impression, please select what kind of emotion (affect) you think the robot was feeling from the following given choices.___________

A: Positive Affect B: Neutral Affect C: Negative Affect

Experiment Music2: After filling in the last question of Video1, subjects listened to a piece of classical music while being shown a 3-minute sequence of facial expressions of the emotional robot with the opposite valance in Video1 simultaneously.

Responses for Music2: After the synthetic robot emotions, subjects were asked to fill in a copy of Affect Grid before answering the following questions:

1. As a total impression, please select what kind of emotional music you think you were listening to from the following given choices. ___________

A: Happy Music B: Neutral Music C: Sad Music

2. As a total impression, please select what kind of emotion (affect) you think the robot was feeling from the following given choices.___________

A: Positive Affect B: Neutral Affect C: Negative Affect

Appendix Ⅶ: The Procedures of Experiments Music1 and Image2 Conducted in Pairs

Experiments Music1 and Image2 were conducted in pairs. In Music1, a piece of classical music was played at the same time as the sequence of robot expressions. In Image2, a set of selected affective pictures was presented with the robot behaviour simultaneously. The two sequences of facial expressions (positive and negative) of the emotional robot used in the two experiments were the same as those used in experiment Speech1. The affective pictures were as the same as that in experiment Image1. The classical music was as the same as that in experiment Music2 which was about 3-minutes long.

Subjects in the beginning of experiment Music1 filled in a similar questionnaire as provided in the previous experiments. Again, no information sheet was given to each subject as instructions. Being similar to experiment Video1, before the warm up procedure, they read a piece of paper (see Fig. Appendix Ⅶ) as follows:

Fig. Appendix Ⅶ: Experimental Instructions of the two experiments Music1 and Image2

When the participant was reading the above paper, he/she received a verbal description in English (a Chinese version can be provided if a Chinese participant requested) of the whole

procedures from the experimental conductor, e.g., as follows:

Firstly, you will watch six different facial expressions of an emotional robot and six different affective pictures on the computer screen at the same time, just to warm you up. Secondly, you will listen to a piece of classical music and watch a corresponding sequence of facial expressions of the robot simultaneously. Then you will need to fill in a short questionnaire. In the third part of the study, you will watch a film clip and a corresponding sequence of facial expressions of the same robot at the same time. Finally, you will need to fill in a final questionnaire.

Then, he/she went through the following procedures:

Warm up: Six different facial expressions of the emotional robot CIM and six static affective pictures on a computer screen selected from the IAPS were showed simultaneously to all subjects. This was exactly as the same as that in the second experiment.

Experiment Music1: Subjects listened to a piece of classical music while being shown a 3-minute sequence of facial expressions (either positive or negative) of the emotional robot simultaneously.

Responses for Music1: After the robot emotions, subjects were asked to answer the following questions:

1. As a total impression, please select what kind of emotional music you think you were listening to from the following given choices. ___________

A: Happy Music B: Neutral Music C: Sad Music

2. As a total impression, please select what kind of emotion (affect) you think the robot was feeling from the following given choices.___________

A: Positive Affect B: Neutral Affect C: Negative Affect

Experiment Image2: After filling in the above questions, subjects watched the affective pictures while being shown a 3-minute sequence of facial expressions of the emotional robot with the opposite valance in Music1 simultaneously.

Responses for Image2: After the synthetic robot emotions, subjects were asked to answer the following questions:

1. As a total impression, please select what kind of affective pictures you think you were viewing from the following given choices. ___________

A: Pleasant Picture B: Neutral Pictures C: Unpleasant Pictures

2. As a total impression, please select what kind of emotion (affect) you think the robot was feeling from the following given choices.___________

A: Positive Affect B: Neutral Affect C: Negative Affect

Appendix Ⅷ: The Procedures of the Experiments Video3 and Speech2 Conducted in Pairs

Experiments Video3 and Speech2 were conducted in pairs. In Video3, a film clip was presented before the sequence of synthetic robot emotional expressions. In Speech2, a piece of recorded BBC News was presented at the same time as the sequence of robot expressions. The two sequences of facial expressions (positive and negative) of the emotional robot used in the two experiments were the same as those used in experiment Speech1. The film clips were as the same as that in the experiment Video2. The recorded BBC News was as the same as that in experiment Speech1.

Subjects in the beginning of experiment Video3 filled in a similar questionnaire as provided in the previous experiments. Again, no information sheet was given to each subject as instructions. Being similar to experiment Video1, before the warm up procedure, they read a piece of paper (see Fig. Appendix Ⅷ) as follows:

Fig. Appendix Ⅷ: Experimental Instructions of the two experiments Video1 and Speech2

When the participant was reading the above paper, he/she received a verbal description in English (a Chinese version can be provided if a Chinese participant requested) of the whole

procedures from the experimental conductor, e.g., as follows:

Firstly, you will watch six different facial expressions of an emotional robot to warm you up. Secondly, you will watch a film clip before you will watch a corresponding sequence of facial expressions of the robot head simultaneously. Then you will need to fill in a short questionnaire. In the third part of the study, you will listen to a piece of recorded BBC News and watch a corresponding sequence of facial expressions of the same robot at the same time. Finally, you will need to fill in a final questionnaire.

Then, he/she went through the following procedures:

Warm up: Six different facial expressions of the emotional robot CIM were showed all subjects.

Experiment Video3: Subjects watched a film clip before being shown a 3-minute sequence of facial expressions (either positive or negative) of the emotional robot.

Responses for Video3: After the robot emotions, subjects were asked to answer the following questions:

1. As a total impression, please select what kind of emotional film clip you think you was watching from the following given choices. ___________

A: Amused (Happy) Film Clip B: Neutral Film Clip C: Sad Film Clip

2. As a total impression, please select what kind of emotion (affect) you think the robot was feeling from the following given choices.___________

A: Positive Affect B: Neutral Affect C: Negative Affect

Experiment Speech2: After filling in the above questions, subjects listened to a piece of recorded BBC News while being shown a 3-minute sequence of facial expressions of the emotional robot with the opposite valance in Video3 simultaneously.

Responses for Speech2: After the synthetic robot emotions, subjects were asked to answer the following questions:

1. As a total impression, please select what kind of News you think you were listening to from the following given choices. ___________

A: Positive News B: Neutral News C: Negative News

2. As a total impression, please select what kind of emotion (affect) you think the robot was feeling from the following given choices.___________

A: Positive Affect B: Neutral Affect C: Negative Affect

Appendix Ⅸ: The Procedures of the Preliminary Experiment

Subjects in the beginning of experiment will receive a verbal description in English of the whole procedures from the experimental conductor, e.g., as follows:

Firstly, you will be asked to read some Instructions of Affect Grid which will be used later in the study, and to fill in a copy of the Affect Grid to measure you mood state. Secondly, you will be asked to listen to two pieces of music, and after each piece of music you will fill in a copy of the Affect Grid. Thirdly, you will asked to watch four film clips one by one on a computer screen, and after each film clip you will asked to fill in a copy of the Affect Grid, and fill in a short questionnaire to indicate how well you can recognize the film clip. Finally, you will rate and compare the emotional valence of all the four film clips together. ***Note that the six piece of emotional materials used in this experiment are not related to each other.***

Then, each participant will go through the following procedures:

Preparation: After subjects have read the instructions for the Affect Grid, they will be asked to fill in a copy of it to indicate their mood states before they listen to two pieces of music on a computer screen through a stereo headphone.

Experiment: After subjects have listened to the first piece of music, they will be asked to fill in a second copy of the Affect Grid to indicate their current state, before they listen to the second piece of music. And they will be asked to fill in a third copy of the Affect Grid after they have listened to the second piece of music. Then they will watch four film clips one by one on the same computer screen through the same stereo headphone.

Responses:

1. Every time when subjects have watched one film clip, they will be asked to fill in a copy of the Affect Grid before answering the following question:

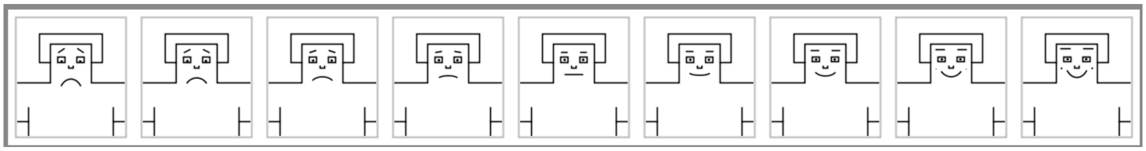
   As a total impression, please select what kind of film clip you think you were viewing from the following given choices. ___________

   A: Positive Film Clip B: Neutral Film Clip C: Negative Film Clip

2. After subjects have filled in the Affect Grid for another four times and answered the above question for four times, they will rate the emotional valence of the four film clips as follows:

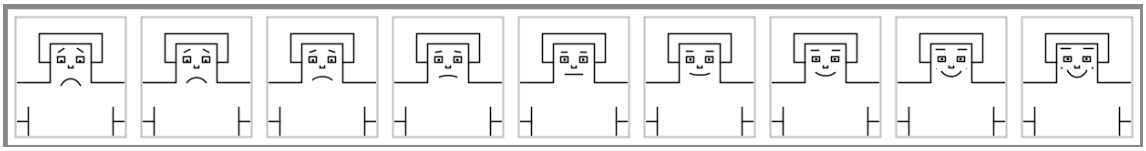
   Now that you have watched all the four film clips, please use the following Self-Assessment-Manikin scales to rate the emotional valence of each of them. Circle any number underneath any of the figures (1 indicates most negative and 9 indicates most positive):

Please circle your answer for the **FIRST** film clip:



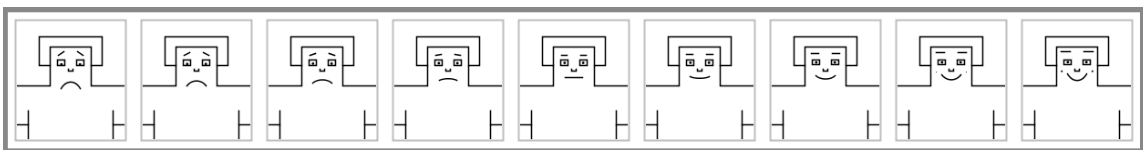
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

Please circle your answer for the **SECOND** film clip:



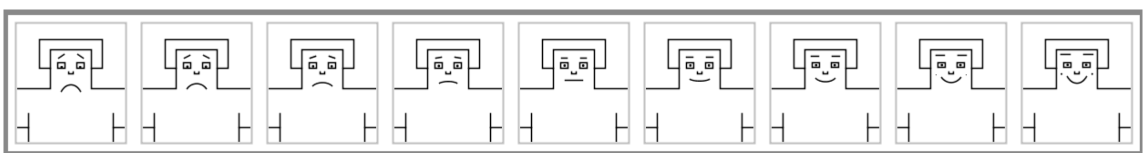
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

Please circle your answer for the **THIRD** film clip:



1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

Please circle your answer for the **FOURTH** film clip:



1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

Appendix Ⅹ: The Questionnaire of Each Proposed New Experiment

An Investigation of How Users Perceive the Facial Expressions of an Emotional Robot

Jiaming Zhang

Dr. Amanda Sharkey, Prof. Noel Sharkey

Neurocomputing and Robotics Group

Department of Computer Science, the University of Sheffield, UK

Information Sheet

1. Research Project Title:

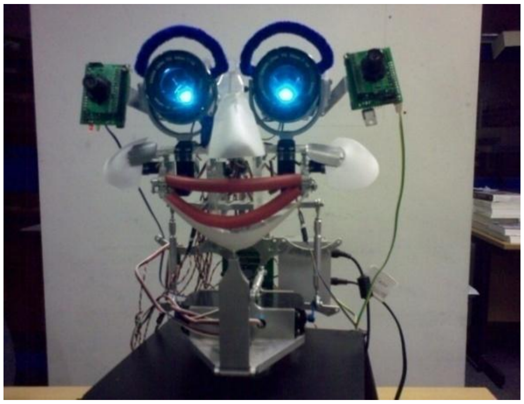
An Investigation of How Users Perceive the Facial Expressions of an Emotional Robot

2. Invitation paragraph

You are being invited to take part in a research project. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

3. What is the project's purpose?

In this project, a mechanical robot head called CIM (see the picture on the next page) in the Neurocomputing and Robotics lab in the University of Sheffield, was programmed to express various simulated emotions, such as Paul Ekman' six universal emotions: happiness, sadness, anger, disgust, surprise, and fear. The main purpose of this project is to find out more about how people perceive the facial expressions of this emotional robot. We are interested in whether the context of a facial expression affects how it is seen.



4. What is the overview of the project procedures?

Firstly, subjects will be asked to read the instructions for the Affect Grid which will be used to measure their mood states later, and to fill in a copy of the Affect Grid. Secondly, subjects will listen to a piece of classical music. After listening to the music, subjects will fill in a second copy of the Affect Grid. Then they will watch a film clip on the computer screen while being shown a sequence of robot expressions at the same time. Finally, they will fill in a third copy of the Affect Grid, and answer two questions to indicate their views of the robot expressions and the film clip. *Note that the music, the film clip, and the sequence of robot expressions used in this experiment are not related to each other.* Further explanation of the project can be given if required at the end.

5. Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep (and be asked to sign a consent form) and you can still withdraw at any time without it affecting any benefits that you are entitled to in any way. You do not have to give a reason.

6. Will my taking part in this project be kept confidential?

All the information that we collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified in any reports or publications.

Thank you very much for taking part in this project!

Date:

Name of Applicant: Jiaming Zhang

Participant Consent Form

Title of Research Project: An Investigation of How Users Perceive the Facial Expressions of an Emotional Robot

Name of Researcher: Jiaming Zhang

Participant Identification Number for this project: **Please initial box**

1. I confirm that I have read and understand the information sheet dated *[ ]* explaining the above research project and I have had the opportunity to ask questions about the project. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason and without there being any negative consequences. In addition, should I not wish to answer any particular question or questions, I am free to decline. ☐

3. I understand that my responses will be kept strictly confidential. I give permission for members of the research team to have access to my anonymised responses. I understand that my name will not be linked with the research materials, and I will not be identified or identifiable in the report or reports that result from the research. ☐

4. I agree for the data collected from me to be used in future research ☐

5. I agree to take part in the above research project. ☐

________________	________________	________________
Name of Participant	Date	Signature
________________	________________	________________
Lead Researcher	Date	Signature

Participant's Information

Male/Female:___________

Age:_________

Nationality:______________

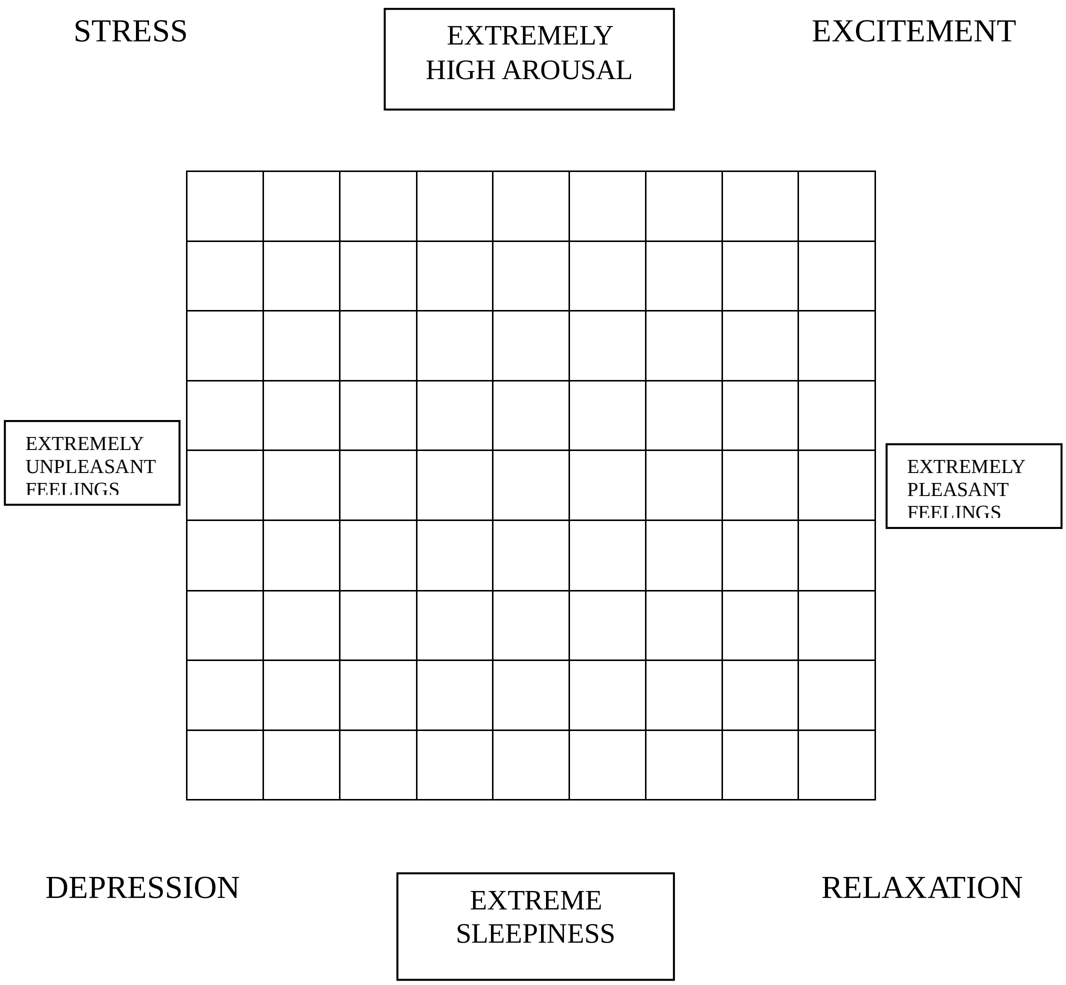
Project Procedures

1. At the beginning of the experiment, you will be asked to read the instructions for the Affect Grid and fill in a copy of it (Page 7).
2. Then you will listen to a piece of classical music played by the computer in front of you through a stereo headphone. After you have listened to the music, please will fill in a second copy of the Affect Grid (Page 8).
3. Please watch a film clip played by the computer through a stereo headphone. At the same time, please also pay enough attention to the robot head when you are watching the film clip.
4. After you have finished watching the robot expressions and the film clip simultaneously, please fill in a third copy of the Affect Grid (Page 9).
5. Finally please answer the last two questions to indicate your views of the film clip, and of the sequence of the robot expressions (Page 10).

Now, please read the instructions for the Affect Grid, and fill in a copy of it on next page.

Affect Grid: Copy One

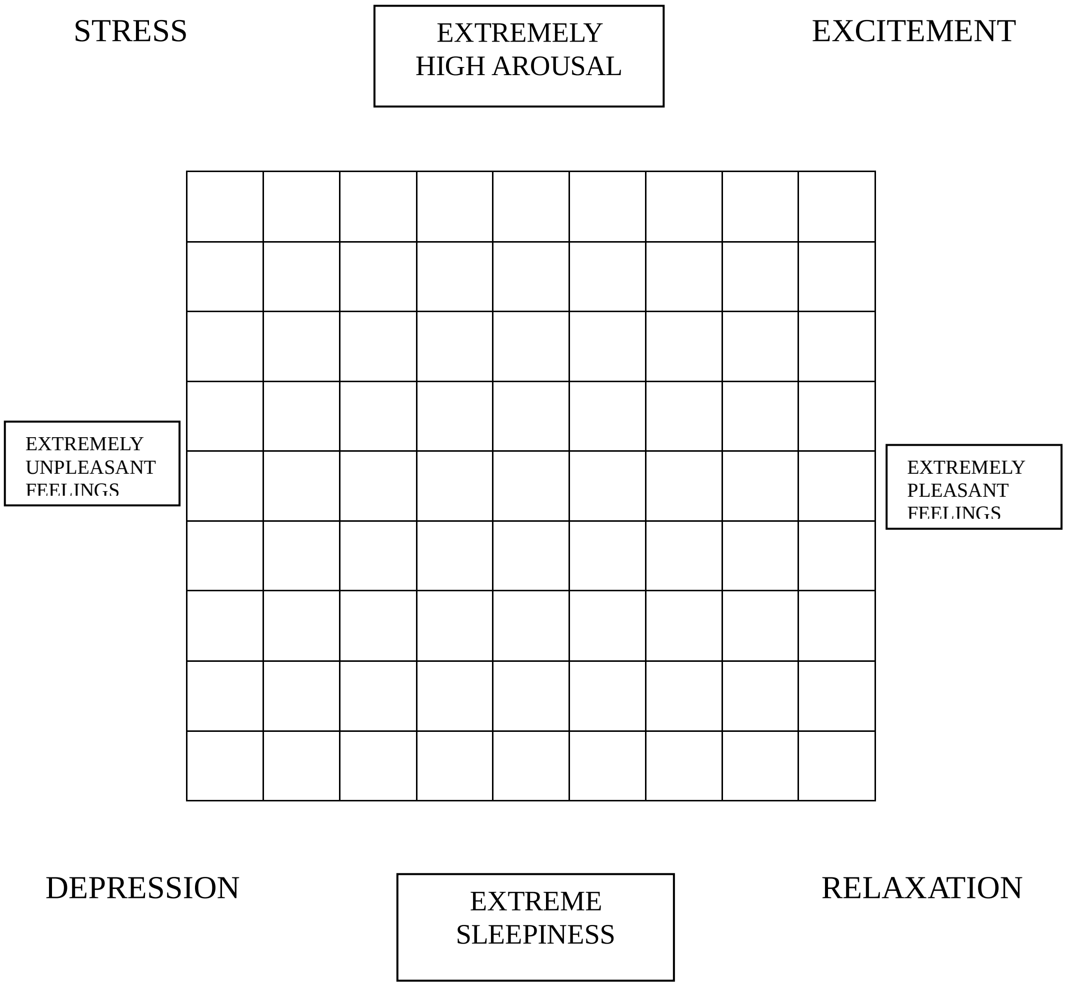
Please rate how you are feeling right now (use X to mark in the grid)



Now, please put on a headphone and listen to a piece of classical music, and then fill in a second copy of the Affect Grid.

Affect Grid: Copy Two

Please rate how you are feeling right now (use X to mark in the grid)



Now, please put on a headphone and watch a film clip on the computer screen while being shown a sequence of robot expressions simultaneously. Please keep in mind that attentions should be paid to both of the film clip and the robot face during the viewing. When you have finished watching them, please fill in the last copy of the Affect Grid and answer the last two questions.

Affect Grid: Copy Three

Please rate how you are feeling right now (use X to mark in the grid)

STRESS

EXTREMELY
HIGH AROUSAL

EXCITEMENT

EXTREMELY
UNPLEASANT
FEELINGS

EXTREMELY
PLEASANT
FEELINGS

DEPRESSION

EXTREME
SLEEPINESS

RELAXATION

Last Two Questions

1. As a total impression, please select what kind of film clip you think you were watching from the following given choices. ___________

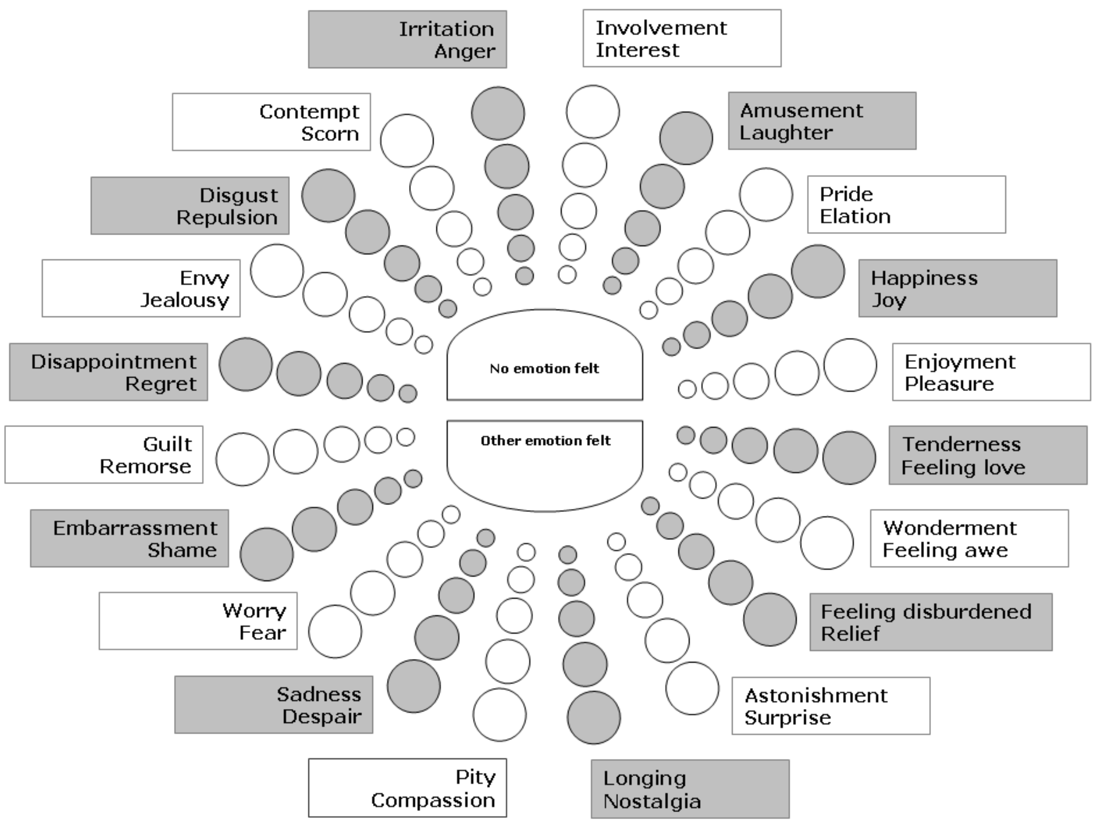
A: Positive Film Clip B: Neutral Film Clip C: Negative Film Clip

2. As a total impression, please select what kind of emotion (affect) you think the robot was expressing from the following given choices.___________

A: Positive Affect B: Neutral Affect C: Negative Affect

Last Two Questions (for the third new experiment only)

1. Please use the Geneva Emotion Wheel (GEW) below to indicate how you felt while watching the film clip. There are altogether 20 emotion families in the GEW. For each item the smallest circle indicates the lowest intensity while the biggest circle indicates the highest intensity. Please choose any emotion family you might felt first and then indicate each emotion family's intensity independently by choosing a circle of that item. If you did not feel any emotion during watching the film clip, please make a cross in the center of the wheel at 'no emotion felt'. If you felt any other emotion during watching the film clip, which is not described by the emotion families presented in the GEW, please write its name in the center of the wheel below 'other emotion felt'.



2. Please use the Geneva Emotion Wheel (GEW) below to indicate what emotions you thought the robot was expressing. There are altogether 20 emotion families in the GEW. For each item the smallest circle indicates the lowest intensity while the biggest circle indicates the highest intensity. Please choose any emotion family you thought the robot might have expressed first and then indicate each emotion family's intensity independently by choosing a circle of that item. If you did not think the robot has expressed any emotion, please make a cross in the center of the wheel at 'no emotion felt'. If you thought the robot has expressed any other emotion, which is not described by the emotion families presented in the GEW, please write its name in the center of the wheel below 'other emotion felt'.