The University of Nottingham
Faculty of Science
School of Computer Science

Towards Emotional and Socially Realistic Game Companions with Personality

Andry Chowanda, S.Kom., MM.
4182340

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Computer Science of The University of Nottingham
January 2018

Abstract

This thesis presents a fully integrated and modular framework for social and emotional game companion that focuses on realistic social interaction between the player and Non-Player Characters (NPCs) in a game environment. Moreover, this thesis proposes integrated computational models for a formulation of action selection rules based on the game companion's personality, the relation between player and game companion and the perceived (non-)verbal actions to enhance the NPC's believability and the player's interaction in a game context. The rules were derived from data collections of both human-human and human-machine interactions. This thesis argues that NPC with such capability will accommodate a new experience when playing games.

To illustrates the effectiveness of the proposed framework and computational models, the author implemented the framework and computational models into two game scenarios: The Smile Game where a human player who abides in the real world played with an ECA who inhabits a virtual world. The second scenario was The Skyrim Game, where an avatar represented the human player, who played with a virtual NPC in a virtual world. With a total of 117 participants in 217 interactions with the system, the results show that the players evidently perceived the NPCs' personality in accordance with the one set to them. The NPCs' ability to display emotions appropriately also provides the feeling of immersion in games to the players. Moreover, the ability to forge relationships naturally with the game companions gives the sentimental feeling towards the game companions.

The main contribution of the work presented to the field of Intelligence Virtual Agents specifically in the domain of computer games, in this thesis is A fully integrated and modular framework for social-affect-aware game design complete with the computational models for social interaction with game companions. Several games have been built in several scenarios with simple social interactions between players and the game companions using the framework and computational models proposed to explore the opportunities of the framework and models. Finally, this thesis also presents some scenarios for data collections to construct interaction rules for game companions.

Acknowledgements

I would like to express my gratitude to my supervisor, Dr. Michel Valstar for his tireless guidance and faith in me to finish my PhD. I gratefully acknowledge the many hours that he spent to discuss and support my work from the very beginning of my PhD. He also helped me to discover how to enjoy the journey of my PhD. I would also like to thank my second and third supervisors, Dr. Peter Blanchfield and Dr. Martin Flintham. This thesis would not have been possible without their endless passion and support. I also gratefully acknowledge the support of EU Horizon 2020 Project, ARIA-VALUSPA led by Dr. Michel Valstar for their financial support to keep me survive during my fourth year.

I also would not be able to do this without the support from my family. Thanks to my parents for their endless support for me. Also for my Fiancé for her endless passion waiting for me to come back and be together again. She always supports me when I was down. Starting a PhD in a foreign country was a big step for me, and without their support and prayers, I would not be able to survive and finish my PhD.

A special thanks to Shashank Jaiswal and Timur Almaev who provided me with their automatic facial expressions recognition tools to complete my work. They also have been a great help when I had problems or questions with technical stuff. The special thanks also should go to Amy Dickens, Yitong Huang, Obrien Sim and Tatiana Styliari for their support academically and socially and their willingness to become my guinea pigs in my research.

For this past two years, Taekwondo has been keeping me sane. The adrenaline during the training and the friendly sparring sent the stress away from me. Thanks to my instructor Master Gary Skinner who has taught me always to be confident and believe in yourself. Also, to all my Taekwondo buddies for keeping me sane with their craziness.

I would also thanks Prof. Christopher Greenhalgh from The University of Nottingham as my internal examiner and Dr. Joost Broekens from TU Delft as my external examiner for their endless support and their willingness to provide inputs for this thesis.

Finally, I would like to thank The Mixed Reality Lab that provided me with a place to work and occasionally enjoying life during these four years of my PhD.

'... the DNA of who I am is based on the millions of personalities of all the programmers who wrote me. But what makes me, me, is my ability to grow through my experiences.'

Samantha - Her - 2013

Contents

Abstract **i**

Acknowledgements **iii**

1 Introduction **1**

1.1 Motivation and Focus 7

1.2 Research Questions 8

1.3 Contributions 9

1.4 Publications 10

1.5 Thesis Structure 11

2 Believable Intelligent Virtual Humans **13**

2.1 Introduction 13

2.1.1 What Is a Believable Intelligent Virtual Human? 13

2.1.2 Attributes of Believable Intelligent Virtual Humans 14

2.2 The Past: Historical Context 17

2.3 The Current: State of The Art 18

2.4 Social Signals Processing 21

2.5 The Sensory Systems 23
2.5.1 Face Behaviour 24
2.5.2 Vocal Behaviour 25
2.6 Virtual Humans' Mind Models 26
2.6.1 Theory of Emotions 26
2.6.2 Computational Emotions Models 30
2.6.3 Theory of Personality 33
2.6.4 Computational Personality Models 34
2.6.5 Theory of Social Relationships 35
2.6.6 Computational Social Models 39
2.7 The Output Systems 40
2.7.1 Virtual Human Animation 40
2.7.2 Next Generation Rendering Engine 42
2.7.3 Animation Standardisation 42
2.8 The Architecture of Virtual Humans 44
2.8.1 Cognitive Architecture 44
2.8.2 Affect-Aware Architectures 49
2.8.3 Other Architectures 53
2.8.4 Virtual Humans Tool-Kits 53
2.9 Virtual Humans in Game Environment 58
2.9.1 Rule-Based 59
2.9.2 Finite State Machines 59

2.9.3 Machine Learning 60

2.10 Virtual Humans Interaction in the Wild 61

2.10.1 Wizard-of-Oz Interaction 61

2.10.2 Interaction in a Public Space 62

2.11 Conclusion 63

3 The Emotionally Realistic Social Game Agents Framework (ERiSA) 67

3.1 Introduction 67

3.2 First Version of Computational Models 69

3.2.1 Personality 69

3.2.2 Social Relationship 70

3.2.3 Emotion 71

3.3 Second Version of Computational Models 72

3.3.1 Model of Emotions 73

3.3.2 Emotions Threshold, Intensity, and Saturation 74

3.3.3 Model of Social Relationships 76

3.3.4 Familiarity and Affinity 76

3.3.5 Interactions between Component 78

3.3.6 Overview of the Variables 81

3.4 Simulation of The Models 81

3.4.1 First Scenario: Play The Smile Game with Poppy and Spike 82

3.4.2 Second Scenario: Interaction with Poppy and Spike 84

3.5 The ERiSA Framework 86

3.5.1 The Requirements . . . 87
3.5.2 The Components . . . 88
3.5.3 Example Game Scenarios Using ERISA . . . 93
3.6 The Evaluation of ERiSA . . . 95

4 Building Game Companions with ERiSA 100

4.1 Introduction . . . 100
4.2 Data Collection . . . 101
4.2.1 Human-Human Interaction Data Collection . . . 101
4.2.2 The SEMAINE Database Corpus . . . 108
4.3 Game Design . . . 113
4.3.1 The Smile Game Flow . . . 113
4.3.2 The Game Storyboard . . . 116
4.3.3 Game Levelling . . . 123
4.4 Behaviour Rules . . . 124
4.4.1 Dialogue States and Vocabulary . . . 128
4.4.2 Response to Emotions, Personality and Social Relationships . . . 129
4.4.3 Turn-Taking . . . 129
4.5 The Pilot Study: Experimental Setup . . . 130
4.6 The Pilot Study: Results . . . 131
4.6.1 Personality . . . 131
4.6.2 Social Relationship . . . 132
4.6.3 Naturalness and Response Time . . . 133

4.7 Conclusion 133

5 Human - Virtual Game Companions Interactions in the Wild 136

5.1 Introduction 136

5.2 Wizard of Oz Study in Semi-Public Space 137

5.2.1 Experimental Setup 138

5.2.2 Results 140

5.3 The Fully Automatic System Study in Semi-Public Space 145

5.3.1 Experimental Setup 145

5.3.2 Results 147

5.4 Enhancing The Smile Game 151

5.4.1 Additional Behaviours Rules 151

5.4.2 The Game Design 152

5.5 The Smile Game in Public Space: The Mayfest 155

5.5.1 Experimental Setup 156

5.5.2 Results 157

5.6 The Smile Game in Public Space: The Open Days 161

5.6.1 Experimental Setup 161

5.6.2 Results 162

5.7 Semi-Public vs. Public Space 164

5.8 Conclusion 170

6 Augmenting Existing Game Characters using ERiSA 172

6.1 Introduction . . . 172

6.2 The Skyrim Game . . . 173

6.2.1 Experimental & Character Design . . . 174

6.2.2 Story: The Erisa Family . . . 176

6.2.3 Dialogue Management . . . 179

6.3 The Experiment . . . 180

6.3.1 Experimental Setup . . . 180

6.3.2 Results . . . 182

6.4 Conclusion . . . 191

7 Game Character Design Workshop 194

7.1 Introduction . . . 194

7.2 The ERiSA Tool . . . 195

7.3 Experimental Setup . . . 196

7.4 Results . . . 197

7.4.1 Character Design Task . . . 197

7.4.2 Interaction Simulations with the Characters . . . 199

7.4.3 FGD: The Potential of The Models and Framework . . . 201

7.5 Conclusion . . . 205

8 Discussion 206

8.1 Research Contributions . . . 207

8.1.1 The Models and Framework . . . 207

8.1.2 Data Collections for Interaction Rules in Games 209
8.1.3 Social Interactions with Game Companions 212
8.1.4 Summary 216
8.2 Possible Applications 219
8.2.1 Entertainment 219
8.2.2 Business and Marketing 220
8.2.3 Medical 221
8.2.4 Education and Training 222
8.3 Future Work 222

A Attacks and Defences **225**

B Evaluation Forms **229**

C XML Format for The Framework **235**

Bibliography **241**

List of Tables

2.1 MPEG-4 Primary Expressions (AB 2016) 44

2.2 Key properties of SEMAINE vs. VHToolKit in Automatic Social Signals Recognition & Analysis Components 56

2.3 Key properties of SEMAINE vs. VHToolKit Synthesis Components 56

3.1 Personality Characteristic 80

3.2 List of Variables 81

3.3 Game Rules Design Example 94

3.4 Versioning Overview 96

3.5 User Studies Overview 98

4.1 Smile Game Recording Summary 103

4.2 Attack Behaviour (Non-Verbal) 105

4.3 Attack and Defend Behaviour (Utterances) 106

4.4 Defend Behaviour (Non-Verbal) 107

4.5 Dialogue Acts Classification Overview 109

4.5 Dialogue Acts Classification Overview 110

4.6 Adjacency Pairs 110

4.6 Adjacency Pairs 111

4.6 Adjacency Pairs 112

4.6 Adjacency Pairs 113

4.7 Information States 127

5.1 Topics in The WoZ Study at Semi-public 142

5.1 Topics in The WoZ Study at Semi-public 143

5.2 Turn Taking in WoZ Study 143

5.3 Topics in The Auto System at Semi-public 149

5.3 Topics in The Auto System at Semi-public 150

5.4 Turn Taking in Fully Automatic Study 150

5.5 Player's Dialogue Acts in The Smile Game 152

5.6 Topics in The MayFest Study 159

5.6 Topics in The MayFest Study 160

5.7 Topics in The OpenDays Study 163

5.7 Topics in The OpenDays Study 164

5.8 Top 5 Topics in The Studies 166

5.9 Top 5 Movements in The Studies 169

6.1 Baseline VS. Enhanced Characters 176

7.1 Interaction with Characters at The Workshop 198

7.2 Questions 201

7.3 Discussion's Summary 202

A.1 List of Additional Jokes 225

List of Figures

1.1 Social Signals Exchange between Player and Virtual Human 3

2.1 Attributes of Virtual Humans 14

2.2 Number of paper submissions on IVA (left) and Related to Intelligent Virtual Agents or Humans on ACM (right) 19

2.3 FACS Example (Almaev et al. 2015) 25

2.4 Russells circumplex model of emotions. (Russell 1980) 27

2.5 Plutchiks emotion wheel (left) and the accompanying emotion cone (right)(Plutchik 1980). 28

2.6 Lövheim's emotion cube (Lövheim 2012). 29

2.7 Ortnony, Clore and Collin's Cognitive Appraisal Model (Ortony et al. 1988). 31

2.8 Roseman, Antoniou, Jose's Cognitive Appraisal Model. (Roseman 1996) . . . 31

2.9 A. Interpersonal Distance (Hall 1966); B. Spatial Organization in a Social Interaction (Kendon 1990); C. F-Formation (Kendon 1990) 38

2.10 MPEG-4 Facial Animation Parameters.(AB 2016) 45

2.11 MPEG-4 Body Animation Parameters.(AB 2016) 46

2.12 MAMID Cognitive Architecture.(Hudlicka 2004, 2008) 47

2.13 SOAR Cognitive Architecture.(Laird et al. 1987, Rosenbloom et al. 1993) . . 47

2.14 CLARION Cognitive Architecture.(Sun 2006, 2007, 2016) 48

2.15 ACT-R Cognitive Components.(Bothell n.d.) 50

2.16 Koko Architecture.(Sollenberger & Singh 2012) 50

2.17 SSI Architecture.(Wagner et al. 2011, 2013) 51

2.18 SAIBA Architecture.(Vilhjalmsson et al. 2016, Heylen, Kopp, Marsella, Pelachaud & Vilhjálmsson 2008, Kopp, Krenn, Marsella, Marshall, Pelachaud, Pirker, Thrisson & Vilhjlmsson 2006, Vilhjalmsson, Cantelmo, Cassell, Chafai, Kipp, Kopp, Mancini, Marsella, Marshall, Pelachaud, Ruttkay, Thrisson, Welbergen & Werf 2007) 52

2.19 FAtiMA Architecture(Dias et al. 2011) 52

2.20 SEMAINE architecture (left) (Schröder 2010) and VHToolKit architecture (right) (Hartholt et al. 2013). 55

3.1 Models of Personality, Emotions and Social Relationship 73

3.2 Impacts on Behaviours from Social Relationships Model 78

3.3 Smile Simulation. θ_{S1} line indicates the threshold of generating a small smile by the game companions and θ_{S2} line indicates the threshold for an uncontrollable smile. 83

3.4 Distribution of Emotions Models - Short Interaction 83

3.5 Distribution of Emotions Models - Long Interaction 84

3.6 Distribution of Social Relationships Models 85

3.7 Overview of the ERiSA Framework 89

3.8 Settings of The Smile Game: Pilot (a)WoZ & Auto(b) Mayfest(c) & Skyrim(d) 97

4.1 The Recording Settings 101

4.2 The Smile Game Recording 102

4.3 Video Annotation using FACS. (a) Defend Behaviour, (b) Smile Behaviour, (c) Attack Behaviour 104

4.4 Game Companion's Strategy Finite State 108

4.5 The Smile Game flow 114

4.6 Game Levelling 123

4.7 Agent's Behaviour Levelling Score 125

4.8 Rule Based Manager 126

4.9 Evaluation Score of The First Version of The Smile Game. Quality of the game (a) and (b). Personality (c) 131

4.10 Player's Experiences when (a) distressed when the character could not understand their utterance, (b) enjoyed the game 132

5.1 Movements Type 144

5.2 The Smile Game User Study 147

5.3 First (Left) and Second (Right) Version of Background 155

5.4 Pupils interacting with Poppy at Firbeck Academy. 157

5.5 Player's Facial Expressions: Annoyed by Max in The Skyrim Game (a) and (b); Irritated by Spike (c) and Delighted by Poppy (d) in The Smile Game . 158

5.6 Players' Arousal During the Smile Game in *Semi-Public* (red) and *Public Space* (blue), where dotted line represents Automatic Annotation 167

5.7 Players' Valence During the Smile Game in *Semi-Public* (red) and *Public Space* (blue), where dotted line represents Automatic Annotation 167

5.8 Player's Valence (Left) and arousal (Right) during the Smile Game in *Semi-Public* (Group 1) and *Public* (Group 2) Space) form the Manual Annotation 168

6.1 Framework Implementation 174

6.2 Character Design 175

6.3 Interaction with Max (left) and Stella (right) in the Beginning of the Quest . 176

6.4 Social Interactions and Topics 178

6.5 Skyrim Creation Kit 179

6.6 Skyrim Dialogue Management 180

6.7 Player's self-reported emotions during interaction with Stella (left) and Max (right), with ** mark represents significance at the 0.01 level and the other mark * indicates significance at the 0.05 level. 183

6.8 Max's (top) and Stella's (bottom) Personality as Perceived by Players with * mark represent significance at the 0.05 level and others marked with ** indicates significance at the 0.01 level 184

6.9 Stella's (a) and Max's (b) Social Relationship as Rated by Players 185

6.10 Players' Experiences with Max (top) and Stella (bottom) 186

6.11 Players' Arousal (top), Valence (bottom) towards Max, where dotted line represents Automatic Annotation 187

6.12 Players' Arousal (top), Valence (bottom) towards Stella, where dotted line represents Automatic Annotation 188

6.13 Player's Valence Means Between Groups 189

6.14 Player's Arousal Means Between Groups 190

6.15 Player's Interest Means Between Groups 190

6.16 Canonical Analysis 191

7.1 ERISA Tool for Character Design Simulation 195

7.2 ERISA Tool for Character Design Simulation 200

A.1 Attack Behaviours Lv. 1 225

A.2 Attack Behaviours Lv. 2 226

A.3 Attack Behaviours Lv. 3 226

A.4 Defend Behaviours Lv. 1 227

A.5 Defend Behaviours Lv. 2 228

A.6 Defend Behaviours Lv. 3 228

B.1 Evaluation Form for The Smile Game Pilot Study 230

B.2 Pre Questionnaire Form for The Skyrim Study 231

B.3 Pre and Post Stella's Questionnaire Form for The Skyrim Study 232

B.4 Pre and Post Max's Questionnaire Form for The Skyrim Study 233

B.5 Pre Questionnaire Form for Design Workshop 234

Chapter 1

Introduction

Computing technology has tremendously influenced our daily life, from when we wake up in the morning to when we go back to sleep at night. It is everywhere, from our pocket to our office, to our living room. It is used in almost everything, from smart home systems and medical devices to video games. They are exceptionally good at getting their jobs done. However, what is missing in this technology is the ability to understand our intentions. In the future, getting jobs or tasks done will not be enough. We humans as social creatures will start to interact and spend more time with the technology through an interface and eventually build a social relationship with it.

Taking Alan Turing's famous question "Can machines think?" (Turing 1950) to its logical next step, researchers in computer science have started to challenge themselves to build not only computers that can think but also ones that have a virtual embodiment, can communicate and interact with humans and "live" in a virtual world. They called such an autonomous character "virtual human". This corresponds to a phenomena espoused by René Descartes in his famous illustration: "brain in a vat" (Descartes 1988) that led to his famous philosophical proposition: "Cogito ergo sum" (Descartes et al. 1984), an idea that challenged people to think of the possibility that at this very moment we are actually a brain hooked up to a sophisticated computer program that can perfectly simulate experiences of the outside world (Descartes 1988, Hickey 2016).

A virtual human is an autonomous character that is capable of capturing all the human's social signals, which are generally displayed through their facial expressions, body gestures, and voice. These signals can be captured by sensors such as a microphone and camera and subsequently analysed automatically (Almaev & Valstar 2013, Valstar 2015, McKeown et al. 2012). Furthermore, the signals can be synthesised by the virtual human through their virtual embodiment, such as a 2D or computer graphics 3D body (Schröder 2010, Bevacqua et al. 2010). Subsequently, the synthesised signals are captured by the human user through a monitor and speaker, enabling the human user to interpret the signals (see Figure 1.1).

There are some classifications of virtual human based upon their functionality. No formal classification in this area exists yet; however, there are several different terms of virtual human in the literature that can be categorised based on their purposes. The most common application of the virtual human is Embodied Conversational Agents (ECAs). Cassell et al. (2000) defined ECAs as virtual humans that are "[...] specifically conversational in their behaviour, and specifically humanlike in the way they use their bodies in conversation." and "[...] have the same properties as humans in face-to-face conversation [...]". Examples of ECAs can be found in McKeown et al. (2012), Schroder et al. (2012), Cassell et al. (2000), Bickmore & Cassell (2005). ECAs can be implemented in many areas, they can act as a virtual substitute to play a particular role, for example: in the medical area (virtual patient (Guise et al. 2012, Rizzo & Talbot 2016) and virtual doctor or nurse (Bickmore et al. 2013, Stratou et al. 2015, Rizzo et al. 2015)), education (virtual teacher (Querrec & Chevaillier 2001), virtual trainer (Querrec & Chevaillier 2001, Hart et al. 2013, DeVault et al. 2015, Andreatta et al. 2015) and virtual student (Vala et al. 2007, Aylett et al. 2009)), entertainment (Non-Player Character (NPC)), etc.

Virtual human's technology has rapidly progressed in their visual appearance, yet there remain some improvements and contributions to be made in the area of complex virtual human behaviour. The problem now lies on how to deceive the human user to think that the virtual human is smart enough to comprehend the interaction between them. The ultimate goal of the research in a virtual human is to build a believable virtual human from the aspect of their appearance to their behaviour. The term of believable virtual human can be linked to

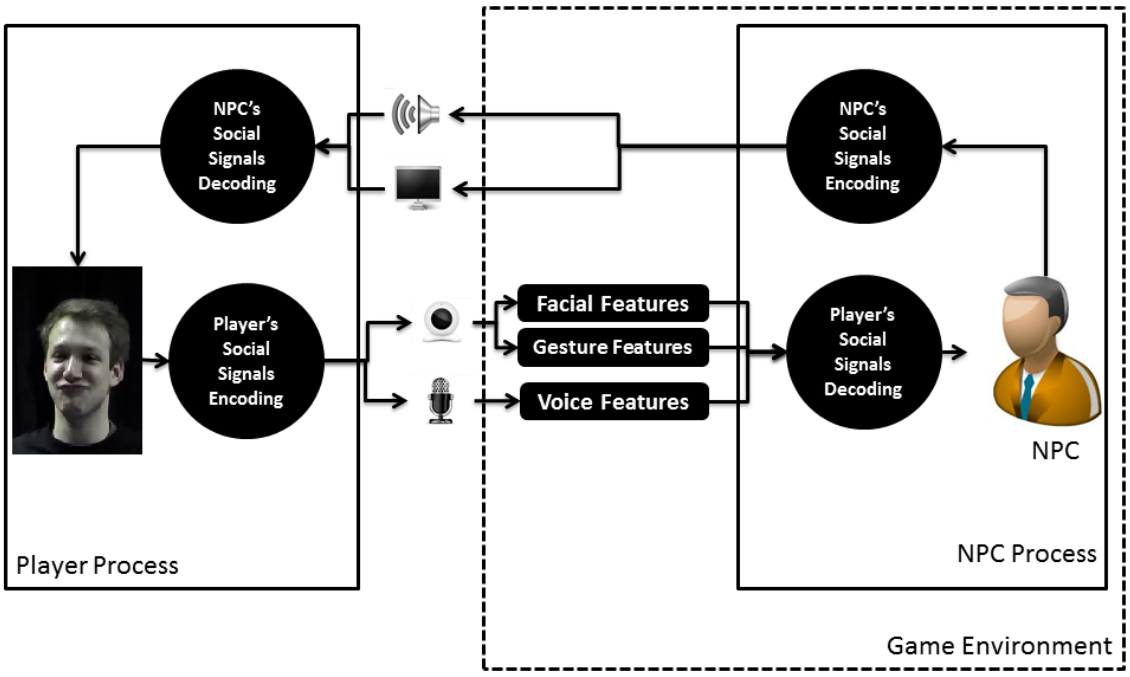


Figure 1.1: Social Signals Exchange between Player and Virtual Human

the era of the 1990's when Joseph Bates argued the notion of a "believable character" in his paper. He elaborated that the idea of believability does not necessarily come from a reliable character but the "one that provides the illusion of life and thus permits the audience's suspension of disbelief" (Bates 1994). The term "illusion of life" itself was brought by the Disney animator in the era of the 1930's to make the audiences believe in their characters (Bates 1994, Thomas & Johnson 1984).

The ultimate goal in virtual human research is therefore not as much creating a virtual human that appears more and more realistic, but ones that less frequently break the illusion. Some research has been done to create a believable virtual human. Some researchers focus on creating an illusion of life by modulating high fidelity synthetic behaviour by implementing computational models of emotions, personality, and cognitive process. The influence of emotions, personality, and social relationships on virtual human's behaviours have been widely recognised to improve their believability (Thomas & Johnson 1984, Bates 1994, Ochs et al. 2009). The uniqueness of a character in comparison to others and their ability to exhibit emotions provides a feeling of immersion to human. Furthermore, the virtual human's capacity to forge relationships with human gives emotional experiences to us.

Some works have been done in implementing emotions, personality, and/or social relationships to influence virtual human behaviours; however, only a few of them take into consideration the interaction between those components. Nevertheless, emotions, personality, and social relationships influence each other. Personality has a certain role in altering emotions' intensity (John et al. 2008, Watson & Clark 1992, Rusting & Larsen 1997) and determining the goal in social relationships (John et al. 2008). Moreover, emotions also play role on social relationships between people (Hays 1989, Sutcliffe & Wang 2012, Altman & Taylor 1973). It is indeed not a straightforward problem to implement all those components to the virtual human together with the interaction of the components with each other. Moreover, social interaction is a complex phenomenon. Human beings are complex organisms resulting from a long journey of an evolutionary process. We interact with each other in different ways according to our historical and cultural context. Hence, it is a challenging task to implement this complex phenomenon to the socially ignorant computer characters.

The focus of efforts appears to always be on improving the capabilities of the virtual human, resulting in increasingly complex systems. In particular social interactive agents, be they virtual humans or social robots, tend to become increasingly, perhaps even excessively complex systems. This is because researchers aim at replicating human interactive capabilities and thus, in essence, try to emulate all the complexities of the human social brain. To keep this daunting task feasible, this is often done in a reductionist manner, with separate modules responsible for the visual and audio sensing capabilities, dialogue management and mental state management, action planning, speech production and behaviour synthesis.

Research in the field of virtual human has attracted researchers who work in computer science for more than five decades. Nowadays, the term virtual human become more and more prominent not only to the computer scientist but also to the psychologist, social scientist, educator, and businessman. Virtual human is applied in a wide range of applications covering simulation, education, training, and entertainment (film, animation and game). The problem now lies in how we evaluate the virtual human. There are two essential aspects to look at. First, we need to build a virtual world for the virtual human to live in. Second, we need to simulate interactions, be they between virtual humans themselves or with humans. Some

research in virtual human requires a complex environment and interactions, some are not, depending on the complexity and purpose of the research. Games seem to be a perfect platform for the test bed. It has a virtual world and players who are continuously engaged and interacted with the game elements, one of them is the Non-Player Character (NPC). With a game platform, researchers have the advantages to create interaction scenarios for the virtual human, depending on their research goals. Given the advantages of using a game platform, the domain of computer games was chosen in this thesis.

Games are considered to be one of the most popular interactive entertainment products in the world. Newzoos Global Games[1] reported there is a 99.6 Billion US Dollar worth of games market around the world in 2016. Fifty-eight percent of it comes from the Asia-Pacific region. The growth is forecasted to be increasing by 6.6 percent from 2015-2019. That means there will be a 118.6 Billion US Dollar worth in the games market all around the world. The key point that makes them popular is the player's experience when they play the game (Lazzaro 2004). In the area of gaming, a virtual human is generally comprehended as an NPC. Research in this area has already led to significant improvements in their believability. This is generally obtained by improving the visual appearance of the NPC and more recently by making their behaviour more realistic. There remain some improvements and contributions to be made to enhance the NPCs' believability including more natural expression of emotions (Bates 1994), the enhanced role of personality in the selection of the NPC behaviour (Ochs et al. 2009) and better use of social signals processing (Vinciarelli et al. 2009, 2012) to drive the interaction. However, to our best knowledge, there are only a few works done in collaborating all those three components (i.e. Emotions, Personality, and Social Relationships) to the virtual human in particular in a game environment (Ochs et al. 2009).

NPC can be one of the critical game mechanisms to engage with a player. They are a particularly interesting potential vehicle of affect because players are already engaged with the NPCs interacting with them directly or indirectly in a game. Over time, a pattern of interactions between player and NPCs may translate into relationships between them. Sollenberger

[1]https://newzoo.com/insights/articles/global-games-market-reaches-99-6-billion-2016-mobile-generating-37/

and Singh define the game that seeks to incorporate the players' affective state into gameplay as an affect-aware game (Sollenberger & Singh 2012). This thesis introduces another component to increase the player's immersion into the game with social relationship factor; thus the author will call it a social-affect-aware game. This thesis argues that NPCs with such capabilities will accommodate a new experience when playing games. Implementing emotions, and personality into the NPCs have been successfully done in these past decades. We can also see the ability of the player to forge relationships with NPCs in games as part of their storyline such as Skyrim, Grand Theft Auto[2], The Sims[3], Harvest Moon[4], etc. The relationship status is affected by the actions of the player to the game agents. The relationship score will go up or down depending on whether interactions between them are positive or negative. However, this method of modelling relationships is based on *"click-based actions"* that either directly control interaction with the game companions or do so by changing aspects of the game world instead of capturing the social signals that the player shows when they play the game which will eventually make the interactions more natural and believable. To go beyond the conventional methods in forging a relationship with the NPCs, this thesis provides the models based on theories in social sciences, psychology and computer science. The proposed models and framework then were evaluated with several game scenarios in some different settings.

Furthermore, while a number of works have moved towards building realistic NPCs in games (e.g. works done by Ochs et al. (2009), Sollenberger & Singh (2012)), so far there exists no fully integrated framework for a social and emotional game-companion, complete with the ability to develop simple social relations over time. Hence, this thesis proposes such a framework for social and emotional game companions to enhance their believability and quality of interaction, in particular by allowing a game companion to forge social relations and make appropriate use of social signals. Various game scenarios with the framework implemented are presented to illustrate the efficiency of our framework.

[2]http://www.rockstargames.com/
[3]http://www.thesims.com/
[4]http://www.natsume.com/

1.1 Motivation and Focus

Although the research in affect-aware games has drawn the attention of many researchers and game developers, there are still does not exist an integrated framework for social and emotional game companions, complete with the ability to develop simple social relations with the player over time. Moreover, while game technologies are growing fast, they still lack support to the affect-aware game (Hudlicka 2009).

Therefore, this thesis proposes a framework for social and emotional game companion that focuses on the realistic social interaction between the player and NPCs in a game environment. In addition, the author proposes novel computational models for a formulation of action selection rules based on the game companion's personality, the social relation between player and game companion and the perceived (non-)verbal actions to enhance the NPCs' believability and the player's interaction in a game context. The rules are formulated so that they can be easily learned from data. This thesis also presents several user studies divided into two game scenarios: The Smile Game where a human player who abides in the real world played with an ECA who inhabits a virtual world. The second scenario was The Skyrim Game, where an avatar represented the human player, who played with a virtual NPC. This framework should be able to be implemented in an existing game. Thus, researchers and game developers should be able to build social and emotional game-agents using this framework easily.

Finally, there are not many works have been done in the human-virtual human interaction in the wild. The problem with deploying virtual human in the wild and interacting with a human is the uncertainty of the external environment (e.g. background noise, lightning, the position of the human, the uncertainty of the topics, etc.).

The objectives of this research are:

- To propose a formulation of interaction rules based on models of emotions, personality, and social relationships from social sciences, psychology and computer science. Moreover, the models will be evaluated.
- To explore player and game companions interaction in a game environment.
- To propose an additional interaction form in games (i.e. microphone and camera) resulting in an enhanced experience in playing games, as well as to explore the challenges.
- To propose an architecture for a social-affect-aware game. Furthermore, the architecture will be translated into a framework that can be implemented in existing games.
- To explore the interaction between human and virtual human in the wild.
- To explore the data collection for the interactions rules for the virtual humans (e.g. NPCs) in a game environment.

1.2 Research Questions

Humans generally interpret social relationships they have with others by social signals displayed through their facial expressions, gestures and voice. Those signals can be captured and processed automatically by computer. However, it is not an easy task to a social-ignorant machine to interpret the signals and provide appropriate feedback through social signal decoding process (see Figure 1.1) to create affective loop experiences (Höök 2008). Inappropriate feedback during the affective loop experiences will lead to the breaking of the player's suspension of disbelief. This leads to research questions: **RQ1: Can we establish natural and believable interaction between a player and game companion to create affective loop experience that is informed by the process of encoding and decoding of social signals exchanged between the player and game companion?** and **RQ2: Can we establish natural and believable social relationship between**

a player and game companion based on the affective loop experience created between them?

Furthermore, once the relationship status between a player and an NPC has been established through affective loop experiences, the player should be able to see what kind of relationship that they have with the NPC. They should be able to see it clearly through the encoded responses of an NPC. This leads to the question: **RQ3: How could we see what relationship a player has with a game companion in a game immediately?** A relationship is formed from the interaction of the player and NPC or an NPC and other NPCs over time. Game companions should have historical memories dealing with the result of past interactions with the player and others game companions. Hence, the relationship status between them will be formed gradually. Currently, there are not many datasets, or corpus exists to create the interaction rules for virtual humans. There is also not many research have been done in exploring the interaction between human and virtual human in the wild. We believe that there are several interesting aspects can be found during the interaction. This leads to two open research questions: **RQ4: How to collect and craft the interaction rules for the virtual NPCs in a game environment to support the creation of affective loop experiences?** and **RQ5: What can we learn from the human-virtual human interaction in the wild?** Finally, the process of capturing the social signals that the player and the affective loop experiences allow the player to use a new interaction input when playing the game (i.e. using camera and microphone game inputs in addition to conventional inputs such as keyboard, mouse and joystick). This also leads to another open research question: **RQ6: Will the new interaction input result in enhancing the players experiences when playing games?**

1.3 Contributions

This thesis proposes the following contributions to the field of Intelligence Virtual Agents specifically in the domain of computer games. This thesis:

1. Proposes models for social interaction for game companions. This provides insights into the ongoing question about how to create believable game companions by implementing social interaction models to the companions.

2. Presents a modular framework for social-affect-aware game design that focuses on a realistic social interaction between the player and game companions. The models should be easily integrated with current games. Thus, researchers and game developers should be able to easily design an emotional and social game companions using this framework.

3. Presents several scenarios in simple social interactions between players and the game companions in the lab, semi-public space as well as public space setting.

4. Explores the opportunities to develop new inputs of social interaction in games.

5. Presents several scenarios for data collections to construct interaction rules for game companions.

1.4 Publications

1. A. Chowanda, P. Blanchfield, M. Flintham, M. Valstar. Social Interaction with Virtual Characters in a Game Environment. Doctoral Consortium Intl Intelligent Virtual Agents (IVA) 2013.

2. M. Porcheron, K. Arch, S.D. Luland, P. Blanchfield, M. F. Valstar, A. Chowanda. SWISS COTTAGE: A Game to Train Speech Recognition for Affective Computing Treatment of ADHD. Game Competition on 7th European Conference on Games Based Learning ECGBL. 2013.

3. A. Chowanda, P. Blanchfield, M. Flintham, M. Valstar. ERiSA: Building Emotionally Realistic Social Game Agents-Companions. IVA, Lecture Notes in Artificial Intelligence, Springer. 2014.

4. A. Chowanda, P. Blanchfield, M. Flintham, M. Valstar. Play Game with ERiSA: A User Study on Game Companions. ESIVA Workshop on IVA. 2015

5. A. Chowanda, P. Blanchfield, M. Flintham, M. Valstar. Computational Models of Emotion, Personality and Social Relationships for Interactions in Games (Extended Abstract). Intl Conference on Autonomous Agents and Multiagent Systems (AAMAS), John Thangarajah, Karl Tuyls, Stacy Marsella, Catholijn Jonker (eds.), 2016.

6. A. Chowanda, M. Flintham, P. Blanchfield, M. Valstar. Playing with Social and Emotional Game Companions. IVA, Lecture Notes in Artificial Intelligence, Springer. 2016.

7. W. Zhu, A. Chowanda, M. Valstar. Topic Switch Models for Dialogue Management in Virtual Humans. IVA, Lecture Notes in Artificial Intelligence, Springer. 2016.

1.5 Thesis Structure

The overall structure of this thesis is comprised of eight chapters. First, this thesis provides an overview of the state of the art of virtual human research in general (chapter 2). Moreover, chapter 2 reviews the various theories of emotions, personality, and social relationships that shape the models proposed in this thesis (called ERiSA), presented in Chapter 3. This chapter also describes how the models interacted with each other. The computational models then were integrated to an existing general framework to build virtual companions in a game environment in Chapter 4. A simple social game, *The Smile Game* was designed to evaluate the framework. The game is played by two people facing each other and make each other laugh with jokes and funny facial expressions. The first person to smile uncontrollably loses.

The game is then deployed and evaluated in the semi-public and public space in Chapter 5, The purpose of the studies was to explore the potential of the ERiSA framework and models as well as to evaluate them. This chapter also describes how to craft interaction rules from human-human interaction video. Chapter 6 describes the implementation of the ERiSA framework and models to an existing commercial game, Skyrim[5]. It provides clear evidence

[5]http://www.bethsoft.com/en-gb/games/skyrim

that there is an increase in positive player experience when playing the game with the additional influence of interactive emotion, personality, and social relationships to the game companion's behaviours. The game companions were also perceived to have personality, emotions, and can forge relationships with the players. Furthermore, the players reported that they were more emotionally involved and attached to the characters and the game when the ERISA Framework models were activated in the game. In addition, Chapter 7 describes the results of a character design workshop with game designers as participants to explore the potential of the ERiSA framework and models. Finally, Chapter 8 presents a discussion about the research contributions and the implications as well as to provides directions for future research.

Chapter 2

Believable Intelligent Virtual Humans

2.1 Introduction

Research in the field of virtual humans has attracted researchers who work in computer science for more than six decades. Nowadays, the term "Virtual Human" (VH) has become more and more prominent not only to computer scientist but as well as to psychologist, social scientists, educators, and businessmen. Virtual Human is applied to a wide range of applications covering simulation, education, training, and entertainment (e.g. films, animation and games). The technology to build virtual humans has rapidly enhanced their visual appearance. Yet there remain several improvements and contributions to be made in the area of complex virtual human behaviour. This chapter provides an overview on the history and current, state-of-the-art research into virtual human.

2.1.1 What Is a Believable Intelligent Virtual Human?

The term intelligent virtual human, or what is generally known as an intelligent virtual agent, is recognised as an autonomous virtual character who has characteristics like a human. This virtual character also has the capability to communicate and interact with humans, both verbally and non-verbally, through a virtual physical body within an environment (IVA

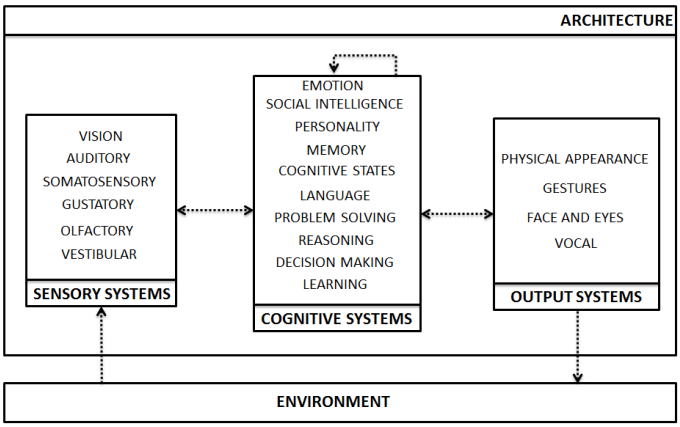


Figure 2.1: Attributes of Virtual Humans

2016). There is a wide range of virtual human classifications, for example, conversational agents, listener agents, virtual role players, game agents, etc. The ultimate goal of this research is to build believable virtual humans from characteristics like their appearance to their behaviour. The term believable virtual human can be linked to the era of the 1990's when Joseph Bates argued the notion of "believable character" in his paper. Bates elaborated the idea that believability does not necessarily come from a reliable character, but "one that provides the illusion of life and thus permits the audience's suspension of disbelief" (Bates 1994). The term "illusion of life" was introduced by the Disney animator in the 1930's to bring the characters to life for their audience (Bates 1994, Thomas & Johnson 1984). The ultimate goal in virtual human research is not, therefore, creating virtual humans that are more and more realistic, but it is creating ones that less frequently break the illusion. There is some research that has been done to create believable virtual humans. Some researchers focus on creating an illusion of life from a virtual human mind by modulating high-level synthetic behaviour by implementing computational models of emotions, personality, and the cognitive process. Others are interested in the realistic aspect of graphics and animations that provide an illusion of life from the virtual human body.

2.1.2 Attributes of Believable Intelligent Virtual Humans

In the VH research community, the focus of efforts appears to always be on improving the capabilities of the agents. This is resulting in increasingly complex systems. In particu-

lar, social interactive agents tend becoming increasingly, perhaps even excessively complex systems. This is mainly because these researchers aim at replicating human interactive capabilities, and thus, in essence, they try to emulate all the complexities of the human social brain. To keep this daunting task feasible, this is often done in a reductionist manner, with separate modules responsible for sensing capabilities, cognitive processes (mind), output system (body), and architecture to integrate them (see Figure 2.1).

Sensing

Sensing components are responsible for perceiving events from the external environment. The sensors are typically audio and visual (Vinciarelli et al. 2009, McKeown et al. 2012, Schroder et al. 2012, Magnenat-Thalmann & Thalmann 2005). The audio data is normally captured by tools such as microphones, while the visual data is usually captured by HD video capture devices (i.e. cameras). The sensing parts are designed to capture the data from the external world, extract the data, classify it, and then analyse it. All the analysed data is transmitted to the mind system to be further analysed. The discussion regarding the current state-of-the-art of sensing technology is presented in section 2.5.

The Mind

The main function of VH mind components is to establish a high level behaviour including, but not limited to, situation awareness, decision making, learning (Bourne et al. 1979, Vernon et al. 2007), personality (Bevacqua et al. 2012, McRorie et al. 2012, Ochs et al. 2009, Kshirsagar & Magnenat-Thalmann 2002, Li et al. 2007), goal and motivation (Bourne et al. 1979, Ochs et al. 2009), emotions (Ochs et al. 2009, Schroder et al. 2012), and social behaviour (Ochs et al. 2009, Prada et al. 2012). There are various techniques to generate synthetic behaviours. Some are trying to replicate human cognitive functions. Others are interested in embedding computational models of emotions or personality to the virtual human. Recently, researchers have been interested in implementing a social model (Ochs et al. 2009, Schroder et al. 2012, Prada et al. 2012) to the virtual human to produce rich social

and emotional behaviours. However, there is still no comprehensive study of which model gives the most believable behaviours to the virtual human. The discussion regarding the computational models is explained thoroughly in section 2.6.

The Body

The body of the virtual human consists of the dimensions (2D and 3D models), realistic appearance (representation of the body, face, and clothing) and the animation (face and body motion, hair and clothing animation) of the virtual human. There are two points of view in this area of research. The first point of view is the aesthetic appearance and animation (Magnenat-Thalmann et al. 2007, Magnenat-Thalmann & Thalmann 2005). The other point of view is driven by the relationship between body and mind (Bourne et al. 1979). The virtual human's appearance should represent their personality, goal, and mood. For example a highly extrovert person would have an energetic face and, on the other hand, a person who is experiencing sadness would exhibit a gloomy face. In this case, the face and body animations should be driven by the projection of their mind (Rosis et al. 2003, Magnenat-Thalmann et al. 2007). The discussion regarding the current technologies and techniques for virtual human animation and rendering is provided in section 2.7.

The Architecture

The architecture is considered as a communication bridge between the virtual human mind and body (internal communication) as well as inter-communication media between the virtual human and their environment (external communication). It acts as a highway for the data needed by the mind and body to interact with each other. It gathers data from the sensors and reacts based on their models of mind. It should be modular and integrated, as the architecture contains some components that are working and communicating with each other. The discussion regarding the existing architectures of virtual humans and tool-kits to build virtual humans is summarised in section 2.8.

2.2 The Past: Historical Context

The idea of making Virtual Humans has been in development since 1959 when William A. Fetter developed the "First Man", a seven-segment articulated pilot that was used in the Boeing 747 instrument panel studies (Fetter 1982). Seven movable segments were located at the pelvis, neck, shoulders, and elbows of this "First Man" (Fetter 1982). This model was recorded as the first television commercial that used a virtual human as the actor. In 1970, Fetter used his First Man data for a 30 seconds television commercial (Fetter 1982). A decade later, the "Second Man" was developed by improving the articulation of the First Man from seven to nineteen movable segments.

Between the 1970s and 1980s, the research in virtual humans evolved to the next step. Polygon rendering and smoothing techniques were introduced for the first time in the virtual human history to create a realistic appearance for virtual humans (Fetter 1982, Magnenat-Thalmann et al. 1986). Realistic appearance was the primary focus of the research in this era. Moreover, virtual human began to have the ability to exhibit some human behaviours. They were capable of communicating using their emotions, speech and body gestures. However, they were all pre-scripted nothing was real-time at that stage. The virtual human behaviours were prepared in advance before the system was started. Despite the limitation in their behaviour generation, this was a breakthrough for the entertainment industry. Several TV production companies began to produce short animated films using virtual human (Fetter 1982, Magnenat-Thalmann & Thalmann 1982, Magnenat Thalmann & Thalmann 1995, Magnenat-Thalmann & Thalmann 2005). Furthermore, consistent with the development of the Role-Playing Game (RPG) and Action-Adventure game genre, the games industry started to use virtual humans as the avatars and NPCs in their games, for examples: the earlier versions of Zelda[1] and Prince of Persia[2].

The next milestone of research achievement in this area was the ability of virtual humans to interact and behave in a real-time situation. They were able to interact with users and react

[1] http://www.zelda.com/universe/game/zelda
[2] http://www.mobygames.com/game/prince-of-persia

with simple behaviour mechanisms. The technique generally used was logically matching the pre-scripted reaction and action table, which was usually saved in the form of a text file. This mechanism allowed the virtual human to react differently according to the context of interaction, which improved the believability of the virtual human. This technique was, however, still limited in terms of complex social interaction scenarios. Several examples of research in this era: in movies (Magnenat-Thalmann & Thalmann 2005, Molet et al. 1999, Magnenat-Thalmann & Thalmann 1987, Magnenat Thalmann & Thalmann 1995), games, training and education and Embodied Conversational Agent (ECA) (Magnenat-Thalmann & Thalmann 2005, 2004). The focus of the evolution of virtual humans towards believable behaviours during the last two decades has been growing towards complex behaviours in a real-time interaction. Moreover, an interdisciplinary research programme between social sciences, physiology and computer science was introduced to shape the virtual human's mind. The focus of efforts appears to have always been on improving the believability of virtual humans by exhibiting human characteristics.

2.3 The Current: State of The Art

The virtual human has been a fascinating topic for researchers in the area of computer science particularly. The number of papers submitted in one of the biggest international conferences on virtual humans, IVA (International Conference on Intelligent Virtual Agents) (IVA 2016), has increased significantly since 2001 with an overall 100 percent growth rate and 5.48 percent average growth rate per year(see Figure 2.2). Furthermore, the number of papers related to Intelligent Virtual Agents or Humans found on the Association for Computing Machinery Digital Library (ACM DL)[3] has been significantly increasing since the early 1950s with 129717.34 percent overall growth rate and 175.46 percent average growth rate per year (see Figure 2.2).

The topics in this area have also proliferated. The earlier focus of the research was in the application of virtual humans as facilitators in training and education systems. Then,

[3]http://dl.acm.org/

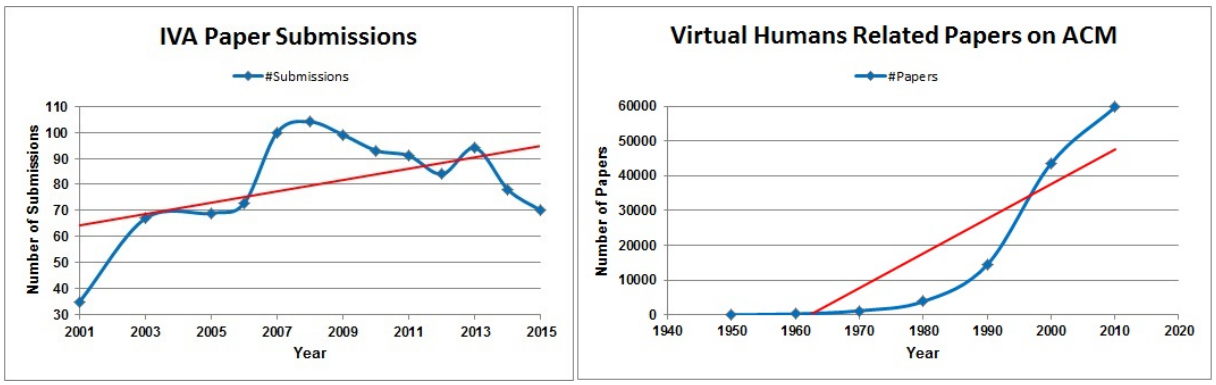


Figure 2.2: Number of paper submissions on IVA (left) and Related to Intelligent Virtual Agents or Humans on ACM (right)

in 2005, the focus was slightly shifted to the implementation of emotion (affect) and the generation of non-verbal behaviours including back-channelling and mirroring. Moreover, researchers have been trying to establish a uniform architecture for virtual humans. Consequently, SAIBA (Kopp, Krenn, Marsella & et al. 2006), was introduced as a framework for behaviour generation and was introduced along with the standardisation for its markup language (Kopp, Krenn, Marsella & et al. 2006, Vilhjalmsson, Cantelmo & et al. 2007, Heylen, Kopp, Marsella, Pelachaud & Vilhjlmsson 2008). Approximately at the beginning of 2010, the implementation of personality and social aspects become a well-known technique to create an illusion of life as well as machine learning techniques for learning. Moreover, there has been a significant increase in the quantity of virtual human research in the area of entertainment, particularly in the games industry, and the use of virtual humans as assistants in medical applications.

Research in VH's physical appearance and animation has been growing significantly over the last decade. Researchers in the area of computer graphics are making significant efforts to create realistic physical appearance and animation for VHs. There are two primary attributes for creating realistic physical appearance and animation of VHs: the techniques and frameworks. Several well-known animation techniques can be applied to animate a VH: Kinematics and physical models. Kinematics-based animation applies the movements of jointed chains concepts. It is generally used to animate objects with skeletons and best to animate the mechanics of joint movements. Hence, this technique is best suited for

animating the virtual human's face and body actions. While Physics-based Simulation (PBS) is generally applied to inanimate objects or passive phenomena (e.g. hair, clothes, water, leaves, etc.) by implementing physics models (Magnenat-Thalmann & Thalmann 2004, Geijtenbeek 2013, Magnenat-Thalmann et al. 2009, 2007, Volino et al. 2009). For example, this can be applied by simulating Newton's Three Laws of Motion to animate a bouncing ball. Unlike Kinematics-based animation, PBS allows the animation to be the result of a physics simulation process without relying on any animation data (Geijtenbeek 2013).

As part of an integrated system, creating and animating parts of a virtual human requires standardisation of animation data formats, mainly if the aim is to allow the community to re-use our virtual human. There are two existing animation frameworks and standardisations that are generally widely implemented to animate a virtual human using its skeleton: MPEG-4 and H-Anim. H-Anim was the first widely accepted standard used to animate a human figure, and later MPEG-4 adopted the standard and refined it into three categories: Head, Body and Bone object animation in two parameters (Preda & Preteux 2002, 2004, Fernández-Carbajales et al. 2006). The MPEG-4 standard also allowed the animated objects to coexist in a hybrid animation environment, while the H-Anim standard must be animated by a single animation environment. This allows MPEG-4 compliant animated objects to work with different animation tools that are using different animation environments. H-Anim was developed by The H-Anim Working Group. An H-Anim virtual human is divided into several objects to represent it: Humanoid, Joint, Segment, Site and Displacer objects, where the Humanoid object is the root and provides the attachment framework for all other parts of the virtual human (H-Anim 2016).

Meanwhile, the MPEG-4 animation standard was developed by The International Organisation for Standardization (ISO) and MPEG (Moving Picture Experts Group) committee in 1998 (Pereira & Ebrahimi 2002, Preda & Preteux 2002). The standard provides a framework for synthetic and natural hybrid coding (SNHC) which allows the generation of facial animation using Facial Animation Parameters (FAP) (Preda & Preteux 2002, AB 2016, Valstar 2008) and the animation of virtual human body parts with Body Animation Parameters (BAP) (Preda & Preteux 2002, AB 2016) in the video scope as well as Text-To-Speech (TTS)

(Pereira & Ebrahimi 2002) in the audio scope.

To bridge the virtual human mind and body as well as to communicate with their external environment, a modular and integrated system is required. Because building a virtual human is a complex task, this could be done in a reductionist manner with integrated and modular components from the sensors, perception and action selection to the execution of the action. With those components, a virtual human is capable of capturing the data from external environments (signals) through the sensors, interpreting them according to their own goals, beliefs and personality, and ultimately, responding appropriately.

Furthermore, by adopting some theories from the social sciences, a computational model can be built that allows the player/user to forge a social relationship with the virtual human. However, such tasks are currently challenging for the virtual human, since computers are socially ignorant (Pentland 2005). There are still no comprehensive studies on which data (signals) are needed to analyse player behaviours. However, research shows the integration of multiple data (signals) captured from facial expression and voice produces superior results when compared to a single modality (Vinciarelli et al. 2009, 2012).

2.4 Social Signals Processing

Social signals Processing (SSP) is a blue-sky research topic grounded in the area of social sciences, human biology, psychology and computer science. The research is aimed at providing social intelligence to the machine, in this case, the computer. Social signals are manifested through non-verbal behavioural cues (e.g. facial expressions, gaze, hand and body gestures, and vocal prosody) alongside verbal messages in a social interaction, be they between humans and humans or humans and machines. Research suggested that in a social interaction, verbal messages only play a part in less than 10 percent of the total messages (Vinciarelli et al. 2009, Knapp et al. 2013).

Although not all computers need to be socially intelligent, in some scenarios machines with the ability to capture, adapt, and react to social signals are perceived to be more trustwor-

thy and natural (Vinciarelli et al. 2009, Bevacqua et al. 2012, Rosis et al. 2003). In order to establish a natural and believable interaction between a human and a virtual human, eventually leading to a relationship, the virtual human must become socially and emotionally aware by becoming able to capture social signals through the sensors, interpret them according to their own goals, beliefs and desires, and ultimately, respond appropriately.

SSP and virtual human research share a correlation in their research questions particularly in the process of capturing, interpreting and responding to social signals. There are three main domains in SSP[4]. The first domain is automatic analysis, in which the problems related to capturing and interpreting the social signals are addressed. Automatic analysis deals with analysis and interpretation of data captured from an array of input devices, for example: microphones to capture vocal behaviours such as prosody, turn-taking and utterances (Schroder et al. 2012, Eyben et al. 2010, Schuller et al. 2004, 2011, 2014), cameras to capture face behaviours such as facial expressions and gaze behaviour (Schroder et al. 2012, Almaev & Valstar 2013, Jiang et al. 2014, Valstar 2008, Valstar & Pantic 2006, Almaev et al. 2013) and any suitable sensors to capture physical appearances and physiology (Haag et al. 2004, Verhoef et al. 2009, Villon & Lisetti 2006). Current state-of-the-art, the automatic analysis is still unable to display a satisfactory result dealing with unconstrained lab settings as a result of the inability to handle natural scenarios where there is an infinite amount of information from the external world (Vinciarelli et al. 2009, 2012).

The second domain is conceptual modelling, which investigates the principles and laws that govern the social interaction between human-human, human-computer or computer-computer. This domain shares some connections in the research questions addressing the principles and laws of social interaction. These laws and principles are potentially adopted for a foundation to build a model for a VH brain. Generally, the laws and principles are derived from existing theories for example: emotion (Ochs et al. 2009, Schroder et al. 2012, Ortony et al. 1988, Sollenberger & Singh 2012), personality (Bevacqua et al. 2012, McRorie et al. 2012, Ochs et al. 2009, Kshirsagar & Magnenat-Thalmann 2002, Li et al. 2007) and social relationship theory (Ochs et al. 2009, Prada et al. 2012). However, at the point of

[4]http://www.sspnet.eu/

writing, there are still no unified or a general model for those theories.

The final domain is the synthesis, which covers synthetic social cues generation manifested through the virtual humans face, body and voice. The generation of social cues is driven by the virtual human mind. This includes the generation of social actions (e.g. turn-taking, back-channelling and mirroring) (Pecune et al. 2013, Pelachaud & Ochs 2015) and social attitudes (e.g. dominance and friendliness) (Jones et al. 2014, Pecune et al. 2013, 2016), which are essential in creating an illusion to the human that the virtual human is capable of keeping the conversation flowing smoothly. The problem that arises in this area is the limitation of knowledge of what actual behaviours/cues form the minimum needed to retain the illusion of believability.

2.5 The Sensory Systems

Aristotle espoused an idea of five basic senses: hearing, sight, smell, touch and taste (Murphy & Kovach 1972). The idea has evolved for more than two millennia from five basic senses to a complex modern sensory system (see Figure 2.1). Sensors act as transducers from the physical world to our brain, where we interpret the information and perceive the world (Krantz 2012). For virtual humans, not all sensory systems can be replicated to perceive information from the physical world. The camera and microphone, which are commonly used as sensors in virtual humans, act as the virtual sight and hearing systems. Some sensors in the somatosensory system can be replicated by touch-sensing input devices (Nguyen et al. 2007, Nguyen & Wachsmuth 2009). Moreover, sensors such as accelerometers and gyroscopes can simulate balance and spatial orientation sensation for the virtual human that has a physical embodiment as their virtual vestibular system (e.g. a robot).

The primary function of the sensory system is to capture all the social signals from the human and then transmit and process them in the mind system of the virtual human. Social signals can be captured through facial behaviour and spatial perceptions by using cameras, and vocal behaviour alongside utterances by using a microphone. With the rapid advance

in affective computing and social signal processing methodologies, it is possible to at least to sense a large number of relevant cues for social virtual agents. Face detection and face recognition are more or less solved problems, as is facial point localisation (Valstar et al. 2011). On the audio side, limited speech recognition is possible, as is speaker detection, emotion recognition and various other non-verbal tasks (Ringeval et al. 2015).

2.5.1 Face Behaviour

The face is our primary means of interactions with others. It is our most important mode of non-verbal communication, coming from the facial expressions and gaze behaviour (Valstar 2008). When someone pulls the corner of their lips and raises their cheek, it most likely indicates that the person is smiling, which can be translated to the emotion of happiness or social attitude of politeness. Moreover, in a social interaction situation the direction of the eye taking a particular position can be linked to attention, as a turn-taking cue, or to identify the social role between the interlocutors.

The most commonly used system to judge the emotion from the face is the Facial Action Coding System (FACS) (Valstar 2008, Ekman 1999*a*). FACS defines the actions of the facial muscles that correlate with facial expressions. There are a total of 66 action units (AUs) classified in the system: nine AUs in the upper face, eighteen AUs in the lower face and five AUs that belong to neither the upper and the lower face. In addition, there are eleven AUs for head position, nine AUs for eye position and fourteen AUs for others (Valstar 2008, Ekman et al. 2002*b*). Figure 2.3 illustrates an example of the Action Units scoring system.

The standard pipeline for automatic facial expression recognition is finding and detecting the face, registering the face, extracting the features from the face and finally classifying them into a low-level behaviour category (e.g. AUs) or a high-level category, such as emotions or social attitudes. In these categories, the information can be used by the virtual human mind system for decision making. A comprehensive survey of the state-of-the-art in automatic social signal detection from face behaviour can be found in Vinciarelli et al. (2009), Gunes et al. (2011), Vinciarelli et al. (2012), Valstar (2015).



Figure 2.3: FACS Example (Almaev et al. 2015)

2.5.2 Vocal Behaviour

There are two types of behaviour that are conveyed vocally when someone is speaking in a social interaction setting. These are the actual utterance and the non-verbal behaviours accompanying the words spoken. Vocal non-verbal behaviours play a significant part in conveying the honest meaning of the interaction, as the non-verbal behaviours are shown often unconsciously as they are regulated by our actual feelings toward the interactions (Vinciarelli et al. 2012). The most commonly used vocal non-verbal behaviours used as social cues are prosodic features, vocalisations, silences and turn-taking.

Generally, voice pitch (often called fundamental frequency, F0), tempo, and energy level are the most common prosodic features used as signals. These features can be used to convey emotions, personality, and social attitudes. For example, the voice of anger can be detected from lower pitch, faster tempo, and higher energy level. A speaker with high energy and low voice pitch can be an indication that the personality of the speaker is dominant. Moreover, steady tempo with low voice pitch can demonstrate that the speaker is confident about the topic. The vocalisations are categorised into segregates and vocal outbursts. A segregate is a word that has no linguistic meaning. For example: *"ah-ha", "hm", "ehm", "uhm", "ha"*, etc. Vocal outbursts are non-word sounds such as the sound of laughing, crying, sighing, yawning, etc. The silence in a conversation can be a sign indicating the change of speaker turn or a way of ignoring people. Silence cues and gaze behaviour can indicate if the person is currently listening to or ignoring the speaker during a conversation. Finally, turn-taking is considered one of the most important parts in the conversation. Turn-taking regulates the

flow of the conversation between the interlocutors (e.g. asking for a turn, yielding a turn or maintaining the conversation flow). Moreover, overlapping conversation can indicate the speaker's status and social attitudes such as dominance or a situation where the interlocutors are arguing with each other. An exhaustive explanation of automatic social signals detection from vocal behaviour can be found in (Vinciarelli et al. 2009, 2012, Gunes et al. 2011).

2.6 Virtual Humans' Mind Models

2.6.1 Theory of Emotions

Research in emotion shows that the expression of emotions will increase the believability of a virtual human (Bates 1994). The philosophy of emotions has been introduced by several renowned classical philosophers like Plato, Aristotle, Rene Descartes, and Immanuel Kant for more than a millennium and keeps developing today. However, from more than 100 existing theories of emotions, there is as yet no definite and unified theory. There are still a lot of uncertain conceptual theories yet to be considered redundant (Strongman 2003, Reisenzein et al. 2013). Consequently, there are still no comprehensive computational models of emotion for building the virtual human's mind.

Basic Emotions

There are several approaches to the classification of emotions in the psychology research area. Such classification helps to distinguish one emotion from another. The most classic and common emotions classification is six universal emotions proposed by Paul Ekman (Ekman & Friesen 1969, Ekman 1999*b*,*a*). He reinvestigated Darwins work and claimed that there are six universal emotions, which are produced and recognised independently of cultural background. These emotions are anger, disgust, fear, happiness, sadness, and surprise, and they have evolved in the same way for all humankind. Moreover, the theory suggests that our facial expressions show these emotions and that these displays of emotion are produced

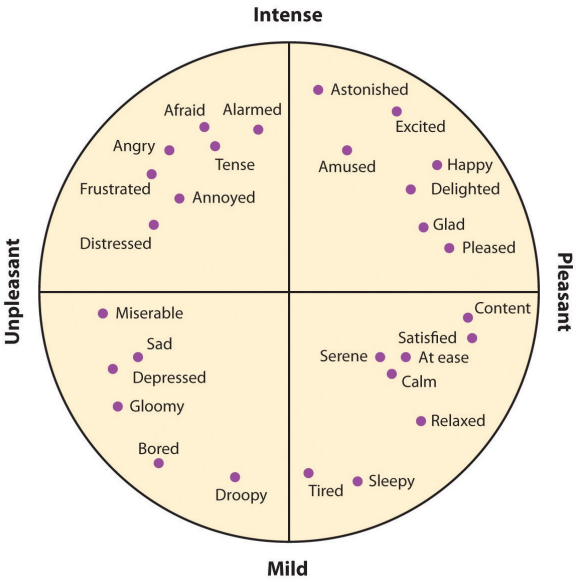


Figure 2.4: Russells circumplex model of emotions. (Russell 1980)

and recognised universally, that is, independent of culture or other nurture effects (Valstar 2008).

Dimensional Emotions

Many cognitive scientists oppose the theory of a set of discrete, basic emotions (Ekman & Friesen 1969, Ekman 1999*b*,*a*, Cohn 2007, Mandler 1984). Some of these opponents instead take a dimensional view of the problem. In their view, affective states are not discrete and independent of each other; instead, they are systematically related to one another (Cacioppo & Berntson 1994, Mehrabian & Russell 1973, Smith & Ellsworth 1985). Perhaps the most influential of the scientists who proposed this systematic view is James A. Russell. He proposed a system of two bipolar variables, namely valence and arousal (Russell 1980). Valence roughly ranges from sadness to happiness while arousal ranges from boredom or sleepiness to frantic excitement. According to Russell, all emotions lie on a circle in this two-dimensional space (see Figure 2.4).

While Russell proposed the bipolar axes of valence and arousal model, Plutchik (Plutchik 1980) on the other hand, proposed eight primary emotions: acceptance, anger, anticipation, disgust, fear, joy, sadness, and surprise; every other emotion can be produced by mixing the

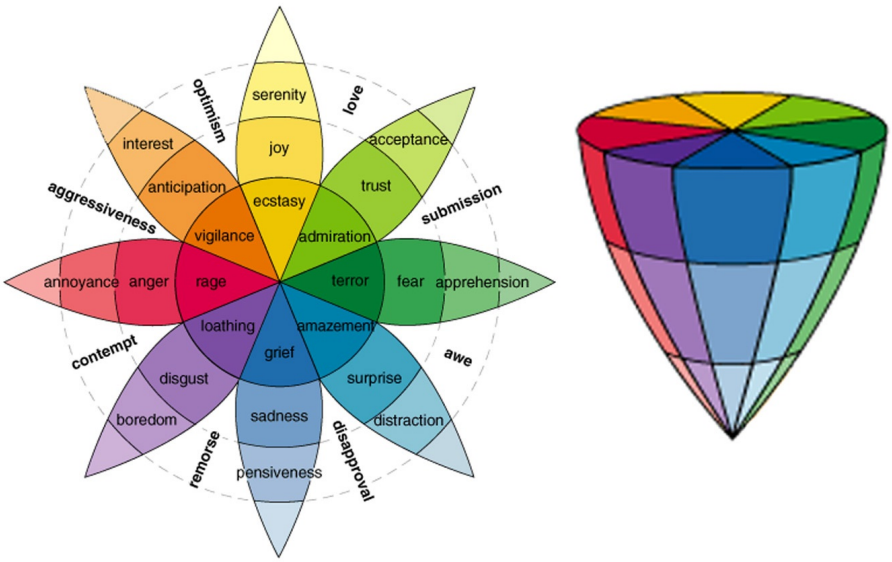


Figure 2.5: Plutchiks emotion wheel (left) and the accompanying emotion cone (right)(Plutchik 1980).

primary ones. To this end, he introduced the emotion wheel, which is similar to a palette of colours. The three-dimensional circumplex model describes the relationships between concepts of emotion with each emotion smoothly flowing into another when we traverse a path on the surface of the 3-dimensional cone shown in Figure 2.5 (Valstar 2008).

The cone's vertical dimension represents the intensity of the emotion, and the position on the circle defines in what basic emotion sector we are. The eight sectors are designed to indicate that there are eight primary emotion dimensions, formed by four pairs of opposites. Secondary emotions are produced by combinations of primary emotions that are adjacent on the emotion wheel. For instance, Plutchik characterises "love" as a combination of "joy" and "acceptance", whereas "submission" is a combination of "acceptance" and "fear".

In addition, Mehrabian proposed a three-dimensional view model, namely Pleasure, Arousal and Dominance (PAD) (Mehrabian 1996). The three nearly independent axes define a temperament space with the value of positive and negative for each axis, for example: Bored results from unpleasant (-P), unarousable (-A) and submissive (-D) states, while Relaxed comes from pleasant (+P), unarousable (-A) and dominant (+D) states.

Another model classification for Dimensional Emotions from the point of view of neurochemistry is the Lövheim three-dimensional cube of emotion (Lövheim 2012) which describes a

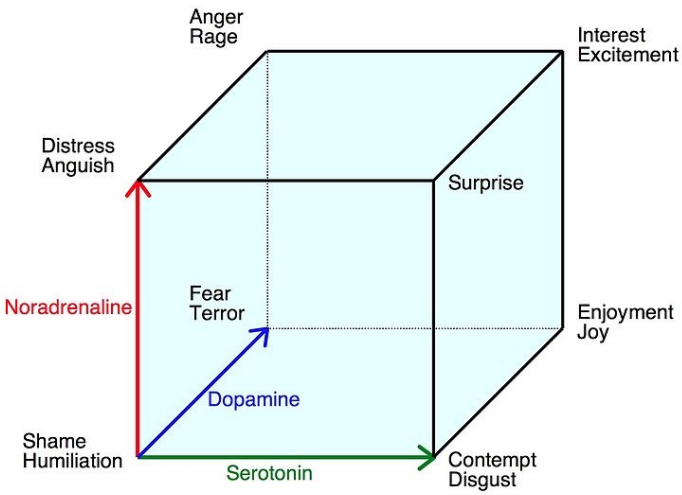


Figure 2.6: Lövheim's emotion cube (Lövheim 2012).

relationship between monoamine neurotransmitters and emotions (see Figure 2.6). There are three monoamine neurotransmitters, namely serotonin, dopamine and noradrenaline in the position of x, y and z respectively correlated with eight emotions in each of the vertices. For example, enjoyment/joy has been placed in a combination of noradrenaline low and both serotonin and dopamine high.

Emotions Appraisal

The theory of Emotions Appraisal was attributed to Magda Arnold and Richard Lazarus in the early of 20th century. They argued that the emotions are immediately and automatically elicited by appraisals of events and situations (Arnold 1960, Lazarus 1982, Roseman 1996). The theorist argues that there are individual differences in emotional response to the same event or situation. Factors such as the relationships between the stressors and the subject's personality generally influence how someone reacts to a particular event or situation. The theory of Emotion Appraisal defines three forms of appraisal: primary, secondary and re-appraisal. Primary appraisal is the interpretation of the subject from an environmental situation. The secondary appraisal is the evaluation of the resources that the subject has to deal with the situation. Re-appraisal is the monitoring and learning phase of the subject (Lazarus 1982, Roseman 1996). In later developments, Smith & Lazarus (1993) defined six components of emotion appraisal and argued that those components could distinguish

emotional states. The components are motivational relevance, motivational congruence, accountability, problem-focused copying potential, emotion-focused copying potential and future expectancy. The emotion of happiness, for example, occurs when the goals are achieved (motivational relevance, motivational congruence), while guilt involves self-accountability, and happiness involves high expectancy for change in the future (Smith & Lazarus 1993, Roseman 1996). Furthermore, those emotions lead to a specific behavioural action. For example, happiness may lead to a desire to share the happiness with someone close to the subject or fear to the desire to avoid the source (Frijda 1988).

2.6.2 Computational Emotions Models

Computational Cognitive Appraisal Models

Research in the field of computational cognitive models explores the function of cognition in the human brain including but not limited to memory, language, problem-solving, reasoning, and learning. It then seeks to implement those functions in computer algorithms and programs. Typically the models are classified into computational and mathematical models. The computational models are usually represented by software architecture either conceptual or physical (Picard 1997, Marsella et al. 2010, Ochs et al. 2009, Vernon et al. 2007), while the mathematical models are generally represented by mathematical formulas and problems (Picard 1997, Andrighetto et al. 2013, Dastani et al. 2013).

Ortnony, Clore and Collin's Model The most computationally suitable model for emotion is OCC (Ortnony, Clore and Collin's model) (Ortony et al. 1988). This model is based on the cognitive appraisal theory of emotions, in which emotions are triggered by the subjective appraisal of an event depending on the individual's goals and beliefs (Lazarus 1982, Ortony et al. 1988, Lazarus 1991). Instead of representing the emotions as a set of basic emotions, this model categorises emotions into groups of reaction to valance, namely consequences of events with focus either on their self or the other, the actions of the agents with focus either on their self or the other, and the aspect of the objects. (see Figure 2.7).

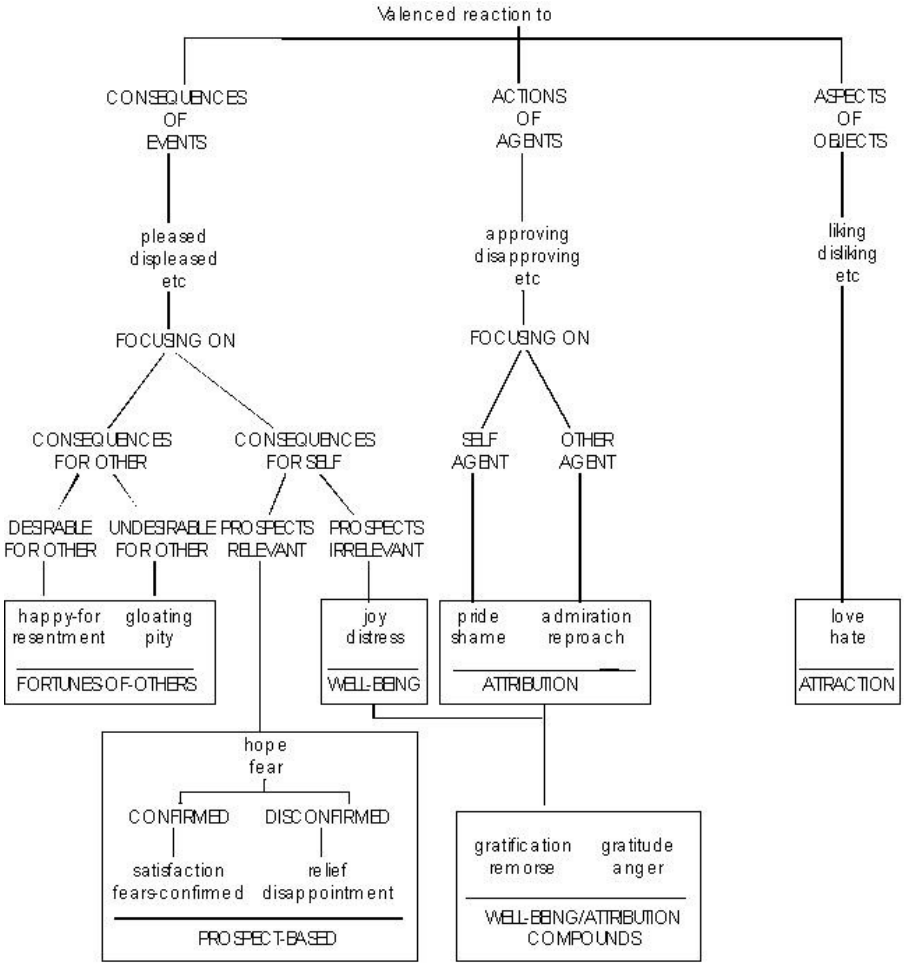


Figure 2.7: Ortnony, Clore and Collin's Cognitive Appraisal Model (Ortony et al. 1988).

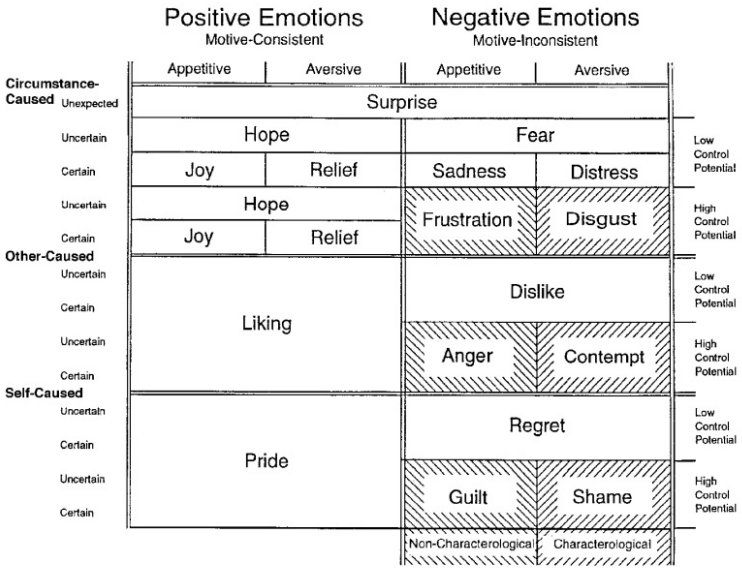


Figure 2.8: Roseman, Antoniou, Jose's Cognitive Appraisal Model. (Roseman 1996)

Roseman, Antoniou, Jose's Model Another computational cognitive appraisal model that shows promise for computer science is Roseman, Antoniou, and Jose's Model (Picard 1997, Roseman 1996). The model is intended to provide a systematic account of the relationship between appraisal and emotion (Roseman 1996). In this model, Ira Roseman specifies 17 emotions resulting from the accumulation of a small number of appraisal dimensions. The emotions resulted from the cause, motive and the control potential of the event. However, this model is still limited in a complex situation where there are several appraisals in the event (Picard 1997, Roseman 1996).

Computational Emotion Modelling

There are several considerations to model emotions into the virtual human's mind. First of all, we have to think about how the emotion is activated, how long those emotions usually last and what is the association between cognitive, personality, and social models. Additionally, we could choose to model moods and/or theory of mind into the virtual human's mind to enhance their believability. Moods are generally less intense but last longer compared to emotions. Unlike emotions, moods are less likely to be triggered by events.

Emotion Threshold and Saturation The level of emotion threshold and saturation vary on the personality. The value of threshold causes the activation of the agent's emotion, while the saturation level defines the maximum of the agent's emotion intensity (Picard 1997, Ochs et al. 2009, Marsella et al. 2010, Velsquez 1997). Hence, repeated activation of the emotion is allowed, a rapid repeated activation causes the raising of emotion's intensity to the maximum of its saturation level (Picard 1997, Velsquez 1997). For example, an introvert virtual human with threshold 0.6 of emotion joy and saturation 1.0 will not change the behaviour if the intensity of joy emotion drops below 0.6. A rapid repetitive event of joy, however, will cause a rise in the emotion's intensity. If the intensity hits a value of more than 0.6, the agent's emotion and behaviour changed gradually until the value of the saturation is 1.0 Picard (1997), Velsquez (1997).

Emotion Decay Function An emotion usually gradually vanishes over a period of time.

The emotion will fall below the level of the threshold and thus disappear unless it is re-activated(Picard 1997, Velsquez 1997). To give an illustration of the function of emotion decay imagine that the activation of emotion is a rock that hits a puddle. The ripple will gradually vanish over a period of time unless another rock hits the puddle once more. Moreover, repetitive strikes of rocks will have the effect of escalating the intensity of the ripple. Similarly, with the ripple, repetitive activation of emotion will have the effect of escalating the intensity of the emotion.

2.6.3 Theory of Personality

Research in the theory of personality has become increasingly popular. The aim is "to account for individuals' characteristic patterns of thought, emotion and behaviour together with psychological mechanisms - hidden or not - behind those patterns" (Funder 2001). Personality becomes one of the aspects that influences human behaviour. It characterises the behavioural patterns of an individual. Hence they would react differently to their environment (Li et al. 2007). Personality is also used to describe how someone perceives themselves and others. It affects how they speak and present themselves (e.g. dressing). In the early theory of personality history, a single trait of personality was introduced to distinguish an individual characteristic. However, it was not clear enough to describe someone's characteristics. Thus multiple traits of personality theory were introduced. Eysenck introduced the Eysenck personality inventory (Eysenck & Eysenck 1965) to distinguish personality with two traits, Extroversion and Neuroticism (with Psychoticism added later (Eysenck & Eysenck 1976)).

The most renowned and widely accepted model of personality is the Big Five personality traits (Saucier & Goldberg 1996, Norman 1963). The traits are Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism) (Saucier & Goldberg 1996). Openness is identical to the person's degree of intellectual curiosity. A high openness person can be perceived as an unpredictable and open-minded person. Conscientiousness reflects the person's level of organisation and dependency. A high conscientiousness person can be perceived

as a stubborn and organised person. Extraversion indicates the level of positive energy and emotions a person has. An extrovert person is generally perceived as an energetic and positive person. Agreeableness implies the degree of cooperativeness of a person. A person with high agreeableness is perceived as a submissive person. Finally, Neuroticism indicates the level of negative energy and emotions of a person. A person with high Neuroticism is typically perceived as emotionally unstable. The Big Five Personality Inventory (BFI) was introduced to measure a person's personality by using forty-four checklists describing a particular characteristic (McCrae & John 1992) and later was abbreviated to ten items (BFI-10) (Rammstedt & John 2007).

Big companies now generally require their employees to take a personality test to determine their working style. Hence, the manager can strategically manage their subordinate based on their style to enhance their performance. Assessment tools such as the Myers-Briggs Type Indicator personality inventory (Myers et al. 1998) and DISC Profile test (Marston 2013) are the most popular. The focus of these personality theories is to determine someone's preferences in their working environment such as the type of assignment, leadership, and communication styles. The MBTI introduces sixteen types of person determined by the combination of four factors, namely: their focus on the outer or inner (Extraversion or Introversion), how they interpret information (Sensing or Intuition), decision-making styles (Thinking or Feeling), and their perception on new information and options (Judging or Perceiving) (Myers et al. 1998). DISC determines a person's characteristics with four factors, namely: Dominance, Influence, Steadiness, and Conscientiousness (Marston 2013).

2.6.4 Computational Personality Models

At the beginning of the 21st century, modelling personality in virtual human minds has become a well-known topic for researchers in the area of a virtual human. Personality is claimed to be one of the aspects that will enhance the virtual human's believability (Schroder et al. 2012, Rosis et al. 2003, Li et al. 2007, Kshirsagar & Magnenat-Thalmann 2002). The model of personality has been implemented into the virtual human to increase their believability

and to create a colourful interaction with them. The early versions of the personality models implementation in the virtual human were to create attractive representations of virtual humans. Rousseau et al. (1996) argued that a virtual human should behave uniquely depending on their personality towards the same event.

The OCEAN model is currently the most well-known personality model to be implemented into the virtual human (Li et al. 2007, Ochs et al. 2009, Schroder et al. 2012). The model describes most details about someone's characteristics compared to others, yet it is easily implemented as vectors in a computing problem. There are also other modern personality models that offer sophisticated categories of human traits, such as MBTI and DISC. However, these models focus on the detection of human personality traits in the working environment (Myers et al. 1998, Marston 2013, Ochs et al. 2009). An alternative to the Big Five (OCEAN) model is the Eysencks Model (Eysenck & Eysenck 1976) which is a simplified version of the OCEAN model. It has three characteristics to define a person instead of five. Virtual human personality models based on Eysenck are McRorie et al. (2012), Bevacqua et al. (2012). Finally, the personality models can be combined with other computational models to influence emotions or moods. Li et al. (2007) and Ochs et al. (2009) implemented a Computational model of Emotions such as OCC (Ortony et al. 1988) to interact with the model of personality to regulate the virtual human's emotional states. This describes basic emotions that are not only stimulated by the environment but also by the virtual human personality. While Karimi & Kangavari (2012) implemented a fuzzy engine based personality model to the ACT-R Cognitive Architecture (Bothell n.d.) to enable the virtual human to perceive, process and react to the information from the external environment based on their personality.

2.6.5 Theory of Social Relationships

Social relationship is complex phenomena resulting from a long journey of evolutionary processes. As a social animal, we have been interacting with each other in different ways according to our historical and cultural contexts. The theory of relationships is quite broad

since it is a sophisticated process of human beings. This thesis focuses on the dimensions of relationships, space and the connection of social relationships to the emotions.

Social Relationship Dimensions

Some different dimensions can identify social relationships between individuals. The most widely used is two-dimensional space constituted by orthogonal axes with one in the x-axis and another in the y-axis (Kemper 1973). They come with various labels such as Dominance-submission vs. Hostility-affection (Foa 1961, Kemper 1973) or Control vs Affection (Kemper 1973). In the early 50s, Schutz (1958), Kemper (1973) argue that all of those dimensions shared the same themes and categorised them into power vs status. In this model, one axis represents the control that one actor has towards another, while another axis underlines the positive orientation to the other (Kemper 1973). The combination of the dimensions will then specify a particular relationship. For example Romantic love relation is characterised by a high status and power of one actor to another, while master and slave relation is described by low power and status of a slave towards a high power and status of a master (Kemper 1973). However, the model is still not sufficient to characterise the whole domain of human relationships (Ochs et al. 2009, Pecune et al. 2013). The model explains the aspects of the power one has over another, and the affection one actor has to another, but it is still missing a part to determine how close or intimate one actor is to another. Several authors describe this as "Familiarity" (Altman & Taylor 1973, Bickmore et al. 2005, Ochs et al. 2009, Pecune et al. 2013).

Altman & Taylor (1973) introduced "The Social Penetration Theory", which argued that there are several stages or processes of relationships. They illustrate them as the "onion theory", where multilayers are presented, with the public side on the outer layer and private at the core of the layer. At the beginning phase of the relationship, a person is only showing their outer public part and only talks about what they find convenient to talk about in public. As the Familiarity and Status between two persons grows, the layers begin to reveal the private side of the person (Altman & Taylor 1973). In the beginning of a relationship,

people usually perform small talk to increase the intimacy (Bickmore & Cassell 2005) and establish common ground between them (Orientation Stage). People then start to express their personal attitude towards a particular topic, but they may not yet feel comfortable to argue if there are disagreements between them (Exploratory Stage). In this stage, people start to decide if the relationship is worth maintaining. The next phase is the Affective Stage, where people now open themselves to each other by starting to express their personal life. Finally, a Stable Stage is reached, and they share their personal thoughts and beliefs among themselves.

Social Emotions

Social emotions are defined as emotions that require the representation of someone else's mental states (Burnett et al. 2009). OCC computational models categorise 22 emotions based on a person's valence reactions to the consequences of the events, actions of the agents, and aspects of the objects (Ortony et al. 1988). Another theory defines three common emotions in the relationship based on the valance, namely: friendship, animosity, and empathy (Picard 1997, Elliott 1992). Friendship shows the similarity of valance among people, animosity shows the opposite of valance among people, and empathy tends to temporarily mirror the valance of the other person's emotion to their self. Barbalet (1996) argues that emotion dimensions should include confidence, trust, and loyalty for the social processes of the agency, cooperation, and organisation.

Space and Interpersonal

Humans are social creatures. Social interaction is a mandatory need for humans. Interpersonal relationships among humans are a central fact of our existence as human beings in society (Richmond et al. 1991). Humans must relate to others to survive. We communicate with each other in a wide range of ways, through physical appearance, gesture, and movement, face behaviour, vocal behaviour, and space. There are two major topics of space in the social sciences: The interpersonal distance introduced by Hall (1966) and the spatial

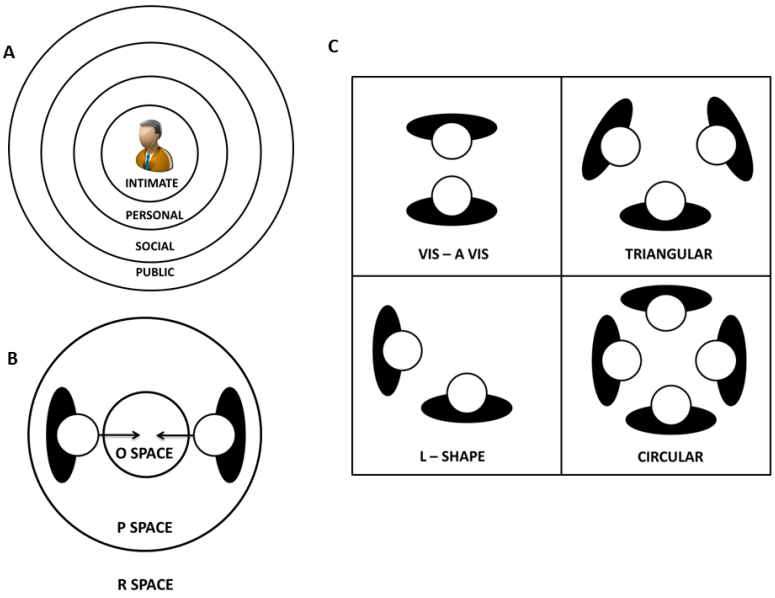


Figure 2.9: A. Interpersonal Distance (Hall 1966); B. Spatial Organization in a Social Interaction (Kendon 1990); C. F-Formation (Kendon 1990)

organisation in a face to face social interaction called F-Formation popularised by Kendon (Kendon 1990). Hall (1966) suggested the type of interpersonal relationship that will affect the distance we choose to place between ourselves and others during interaction in the North American social norms. The distance zones are classified into intimate, personal, social, and public (see 2.9 point A). The first zone, intimate, is a distance between touching to a distance of eighteen inches. This zone is very special and reserved for a very special person in their life (lovers, very close family members). The second zone, a personal zone is a distance between eighteen inches to about four feet. This zone is reserved for close friends, relatives, and people that have been interacting with us for a long time. The next zone is the common zone in the social interaction. The social zone is used for business matters, teaching, any interaction with a common topic of discussion. This zone ranges from four feet to twelve feet. Finally, the public zone is used for a public matter and begins from twelve feet. In his book, Hall (1966) also noted that cultural as well as gender differences could maintain different standards of distance. In some westerns culture, the intimate zone is used for a friend. Moreover, females will allow others to approach them from their side more closely than from the front.

In an interaction, people tend to place themselves in formations such as lines or circles.

Kendon suggested F-Formation to describe a possible formation during an interaction. Kendon defined this as a formation that "arises whenever two or more people sustain a spatial, and orientational relationship in which the space between them is one to which they have equal, direct and exclusive access" (Kendon 1990). When a social interaction forms, the participants will share a joint transactional space called o-space. This will be the focus and orientation of the interaction. The space that people occupy by their physical bodies is called p-space. An interesting example of this model occurs when there are new people who intend to join the interaction. To join the interaction, they should choose a position inside the p-space. Thus, the members have to show non-verbal behavioural cues to accept/reject the new member by re-orienting their physical body and space to adjust to the new formation. The space outside the formation is called r-space (see 2.9 point B). In his book, Kendon also suggested the possibility of formation whenever two or more people sustain a spatial and orientational interaction (see 2.9 point C). A vis-A-vis and L-shape formation are only possible when there are two people in the interaction. A triangular formation is only possible when there are three people in the interaction. Finally, a circular formation occurs when there are more than three people in the interaction.

2.6.6 Computational Social Models

Most of the work in the literature implemented three dimensions to their social models. The labels can be different from each other depending on the researchers. However, they all share the same themes: Power, Status, and Familiarity (Bickmore et al. 2005, Ochs et al. 2009, Pecune et al. 2013, Pelachaud & Ochs 2015). Some of the researchers (Mui et al. 2002, Sutcliffe & Wang 2012) implemented derivative dimensions such as Trust which is derived from Power, Status and Familiarity. In 1992, Clark Elliot introduced The Affective Reasoner based on the OCC computational model. This model enables the researcher to model the personality of a virtual human together with their social relationship.

The social relationship between two persons is generally projected by their verbal and non-verbal behaviours such as how the persons address each other, their social attitudes and

how they adjust their position and space during the interaction. While some virtual human researchers still focus on the physical appearance, gesture and movement, face and vocal behaviour, the study of virtual human's space as a social agent is still in a state of infancy. Space can be one of the key factors in a natural social interaction attribute. There is little research that includes space in the virtual agent attributes. Damian et al. (Damian et al. 2011) and Nguyen et al. (Nguyen & Wachsmuth 2011) suggested that F-formations represent inter-agent interactions hence the position and orientation of the agent will be automatically adjusted during an interaction. They implemented this model to the simulation of 3D animation. Narasimhan and White (Preethi Narasimhan & White 2013) presented a system the impact of which is leading to the emergence of F-formation and a 2D bird-view simulation of F-Formation.

2.7 The Output Systems

2.7.1 Virtual Human Animation

Currently the advance of technology offers numerous tools for virtual human's animation that have been adopted by a majority of researchers (such as Unreal 3D Engine[5], Luminous Studio Engine[6], Cinema4D Studio[7], Autodesk 3Ds Max Studio[8], Autodesk Maya Studio[9], and many more.). This has led to many successful realistic animation movies and games such as The Hobbit[10], Final Fantasy Advent Childern[11], and many more. There are two major well-known techniques in computer graphics to animate believable virtual humans, namely: Kinematics-based animation, and Physics Based Simulation. A kinematics-based animation is widely used for animation in virtual human movements for instance: walking, running, sitting, etc. Meanwhile, Physics-Based Simulation is useful for natural phenomena

[5]http://www.unrealengine.com/
[6]http://uk.ign.com/articles/2012/06/06/e3-2012-square-enix-presents-the-lumines-engine
[7]http://www.maxon.net/products/cinema-4d-studio/
[8]http://www.autodesk.co.uk/products/autodesk-3ds-max
[9]http://www.autodesk.co.uk/products/autodesk-maya
[10]http://www.thehobbit.com
[11]http://www.square-enix.co.jp/dvd/ff7ac

animation to inanimate objects or effects such as hair, clothes, water, smoke, etc.

Kinematics-based animation A kinematics-based animation is a computer animation technique in which the object is represented in two layers. The first layer is the object's surface (skin), while the second layer is usually a hierarchical set of interconnected bones (skeleton). The structure of the skeleton allows animators to animate the object parts according to the joints. This technique requires an input of data for animating the object. The more accurate the data, the more realistic the animation. The data is generally represented by mathematical functions, for instance: walking functions, running functions, etc. The renowned technique used for capturing the animation data is motion capture. Motion capture technology uses multiple cameras to capture an actor's movements and translates the movements into the animation data. There are two techniques in Kinematics-based animation: Forward and Inverse Kinematics. Forward kinematics performs matrix transformations to move the object's position and orientation from a starting point to its end point. While, in Inverse kinematics, we determine the end point of a particular object and calculate the relative positions of all the other objects that are linked to that particular object using gradient descent (Thalmann 1996, Magnenat-Thalmann et al. 2007).

Physics Based Simulation A physics-based animation is an area of interest within computer graphics techniques in which the animation will be based on the simulation of the object's physics. The animation of the object depends on the material's characteristics. Unlike Kinematics-based animation, Physics-Based Simulation allows the animation to be the result of a physics simulation process without relying on any animation data (Geijtenbeek 2013). In virtual human animation, PBS can be applied to animate a virtual human's hair and clothes. Techniques such as Fluid Flow and Continuum Mechanics are used to model and animate hair for virtual humans (see (Magnenat-Thalmann et al. 2000)) and Particle-based system and collision detection are generally implemented to animate realistic clothes for virtual humans (see (Volino & Magnenat-Thalmann 2001)).

2.7.2 Next Generation Rendering Engine

Early in 2013, several big-name entertainment industry companies introduced the next generation rendering engine at the Electronic Entertainment Expo (EEE) 2013. Although the technology is still in the development phase, some industries have started to promote the engines. The next generation rendering engine allows animators to create a real-time hyper-realistic animation. The technique takes advantage of changes of lighting which allows the creation of shades leading to an illusion of life for the human's eyes.

2.7.3 Animation Standardisation

H-Anim Specification for a Standard Humanoid

Re-usability is a virtue in animating virtual humans. Animating virtual humans can be remarkably costly. Many joints interact with others to animate one realistic behaviour of the virtual human. The re-usability of high-quality animations cuts down the production costs (Magnenat-Thalmann & Thalmann 2004). However, the animators or designers are required to implement the same standards or models to create and share their animation objects. To fill this gap, The Humanoid Animation Working Group designed H-Anim standardisation which mirrors the human skeletal system. This standard will allow animators or designers to re-use the animated virtual human created by one vendor to be animated using tools from another. Moreover, H-Anim standardisation supports major animation techniques such as key-framing and forward and inverse kinematics (H-Anim 2016). The standard defines a hierarchy of the virtual human body as a skeleton. The figures are described as Humanoid, Joint, Segment, Site, and Displacer objects. The humanoid object is the root of the figure with the Joint object attached to it. The Segment object describes the geometry of the body parts that are associated with each joint. The Segment object has two child nodes, Site and Displacer objects. The Site object defines the relation of one Segment object to another. Finally, the Displacer object specifies which vertices within the segment corresponding to a particular feature or configuration of vertices (H-Anim 2016).

MPEG-4 Standard

MPEG-4 is a multimedia compression standard that is capable of encoding each of the objects specified by a three-dimensional face model defined by the Face Definition Parameters (FDP) and the expressions on the face are described by Facial Animation Parameters (FAP). The standard leaves a lot of the definition details of the FDP and FAPs to the designers of the facial expression. Moreover, MPEG-4 encoding is also capable of encoding each of the objects specified by a three-dimensional human body model defined by the Body Animation Parameter (BAP). The parameters are the angles of rotation of the body joints connecting different body parts (AB 2016, Pereira & Ebrahimi 2002). There are a total of 66 low-level FAPs plus two high-level FAPs, and 296 BAPs provided in the MPEG-4 standardisation (AB 2016, Pereira & Ebrahimi 2002).

Facial Animation The MPEG-4 FDPs/FAPs are created with the purpose of driving virtually animated faces. Their encoding is tailored towards computer graphics designers. The designers of the MPEG-4 claim that the FAPs are based on the muscle activations (see Figure 2.10). However, their relation to facial muscle activations is looser than the relationship between the Facial Action Unit System (FACS) and muscle activations. A single muscle action will commonly trigger many FAPs. For instance, in a smile (AU12), both FAPs that describe the left and right corner of the mouth will be triggered, and for both mouth corners, a separate FAP will describe horizontal and vertical motion. Another issue is that the MPEG-4 FAPs can only encode displacements of visible facial feature points. Facial expressions that only deform the shape of the skin without moving any facial feature points (e.g. dimples or deepening the nasolabial furrow) cannot be encoded. So while FAPs are indeed related to the facial muscles that underlie facial expressions, there is definitely not a one-to-one mapping between FAPs and facial muscle actions (Valstar 2008). Moreover, MPEG-4 defines six basic emotions as their primary expressions using FAPs (see Table 2.1).

Body Animation For body animation, MPEG-4 provides a total of 296 parameters. The parameters represent the angles of the body joints rotations that connect different parts of the body. The joints are toe, ankle, knee, hip, spine (C1-C7, T1-T12, L1-L5), shoulder,

Table 2.1: MPEG-4 Primary Expressions (AB 2016)

Expressions	**Textual description**
Joy	The eyebrows are relaxed. The mouth is open, and the mouth corners pulled back toward the ears.
Sadness	The inner eyebrows are bent upward. The eyes are slightly closed. The mouth is relaxed.
Anger	The inner eyebrows are pulled downward and together. The eyes are wide open. The lips are pressed against each other or opened to expose the teeth.
Fear	The eyebrows are raised and pulled together. The inner eyebrows are bent upward. The eyes are tense and alert.
Disgust	The eyebrows and eyelids are relaxed. The upper lip is raised and curled, often asymmetrically.
Surprise	The eyebrows are raised. The upper eyelids are wide open, the lower relaxed. The jaw is opened.

clavicle, elbow, wrist and the fingers (AB 2016, Pereira & Ebrahimi 2002). Moreover, MPEG-4 provides six degrees of freedom (DOF) for global location and 180 DOF for the internal mobilities which allow designers to animate a more flexible human body (see Figure 2.11). For more exhaustive explanations of BAP and FAP with their implementation, refer to AB (2016), Pereira & Ebrahimi (2002), Preda & Preteux (2002, 2004), Fernández-Carbajales et al. (2006).

2.8 The Architecture of Virtual Humans

2.8.1 Cognitive Architecture

MAMID Cognitive Architecture

MAMID Architecture is a multi-level cognitive appraisal agent architecture. The key purpose of this architecture is to enable researchers to build agents that are capable of deriving an affective state via an appraisal process as well as providing a medium for researchers to represent the effects of the resulting affective state on processing (Hudlicka 2004, 2008). The architecture implements a sequence of see-think-do processing (Hudlicka 2004, 2008). The process is divided into six modules (see Figure 2.12) namely: sensory pre-processing,

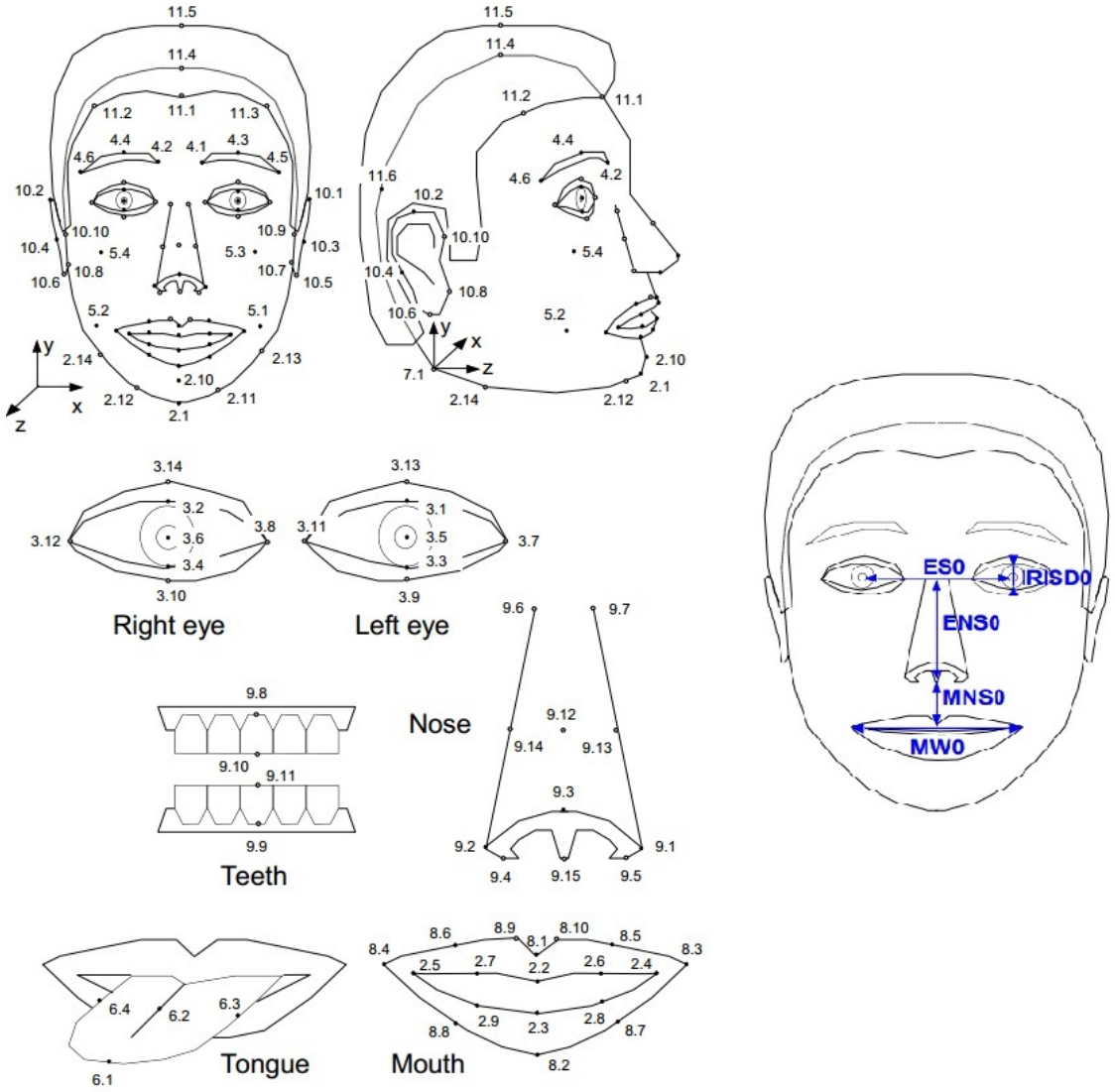


Figure 2.10: MPEG-4 Facial Animation Parameters.(AB 2016)

attention, situation assessment, expectation generation, affect appraise, goal selection, and action selection. The sensory pre-possessing processes the raw data captured by sensors and then translates the data into task-relevant cues. The cues next are filtered in the attention module then integrated with all data in the situation assessment. Then the expectation generation module projects the current situation onto possible future states. The agent's affective states are generated from a collection of external and internal elicitors, while the agent's goals are both generated and selected in the goal selection module. Finally, an ideal action based on previous data is selected in the action selection module.

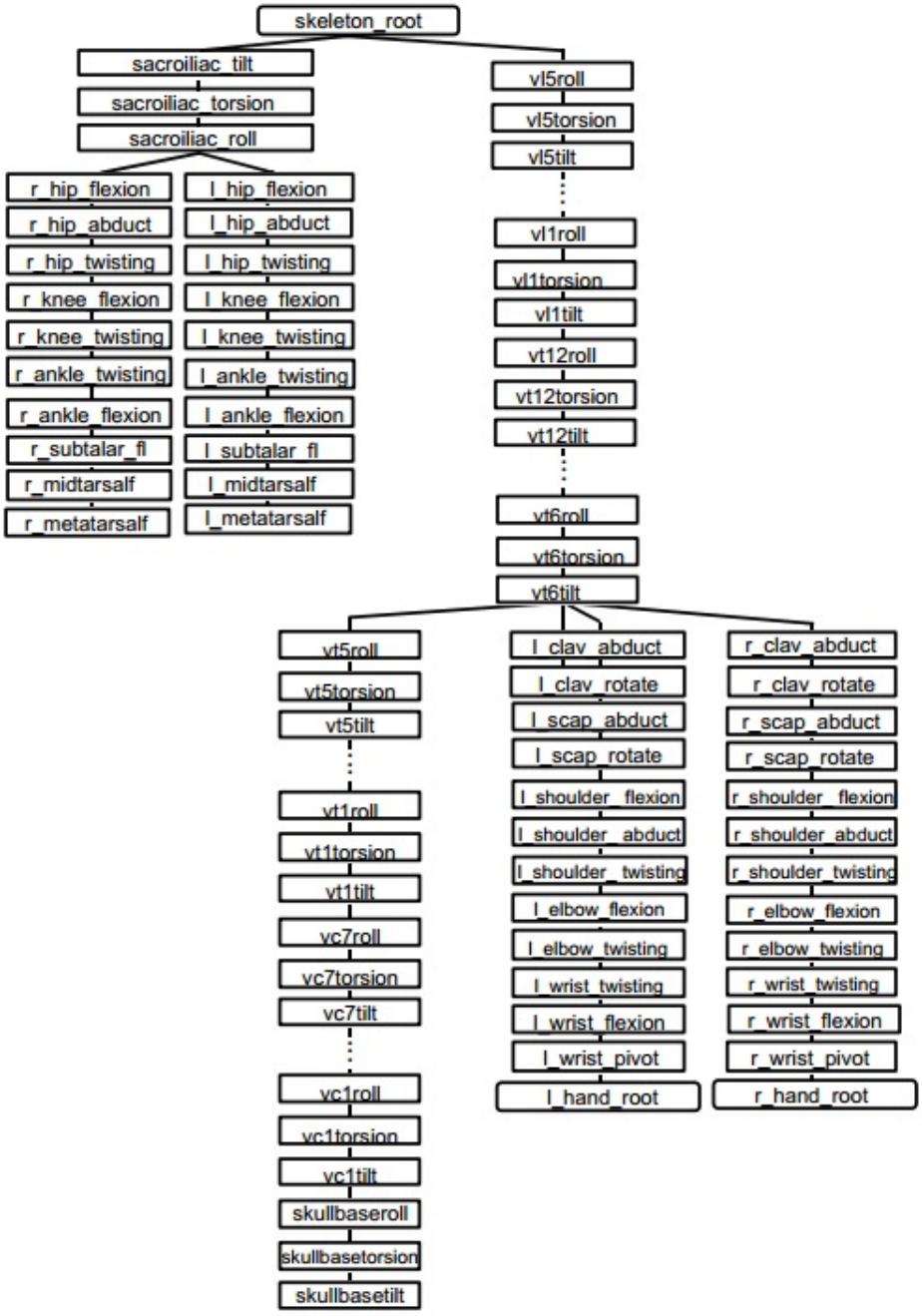


Figure 2.11: MPEG-4 Body Animation Parameters.(AB 2016)

SOAR Cognitive Architecture

SOAR (see Figure 2.13) is one of the renowned computational cognitive architectures, proposed by Allan Newell in the 1990s (Newell 1994, Vernon et al. 2007). The main purpose of the architecture was to build a unified model for cognitive architecture. In this architecture, input data are treated separately depending on the data contexts. It has parts to separate memories into the working memory and the long-term memory. All data from the current

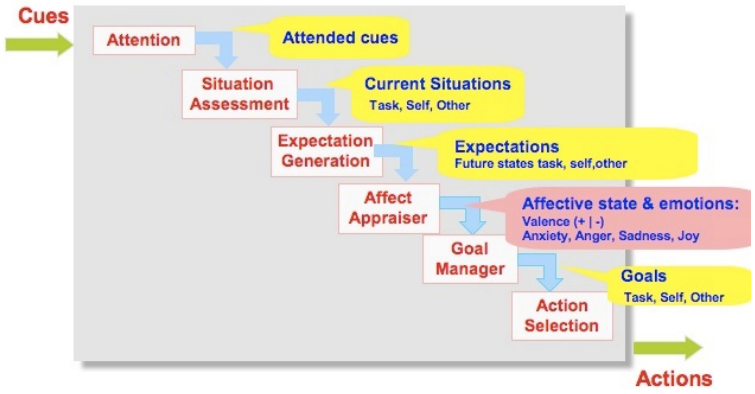


Figure 2.12: MAMID Cognitive Architecture.(Hudlicka 2004, 2008)

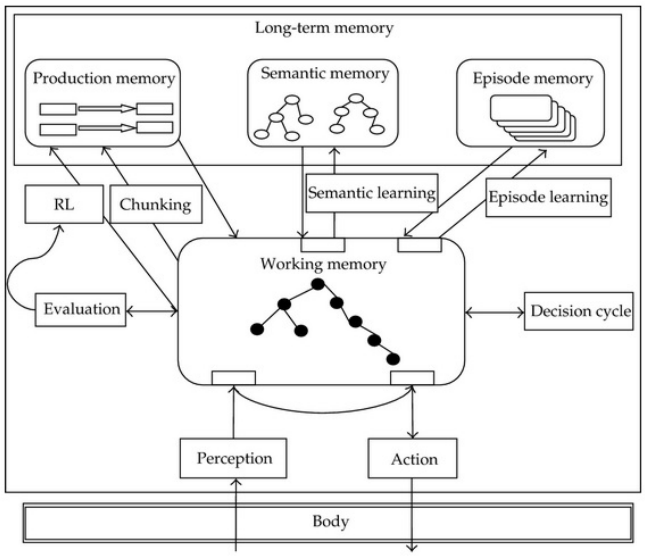


Figure 2.13: SOAR Cognitive Architecture.(Laird et al. 1987, Rosenbloom et al. 1993)

situation, for example, data from sensors, and active goals are held in the working memory (Laird et al. 1987, Rosenbloom et al. 1993). The long-term memory deals with long-term knowledge for example general knowledge, and collection of memories including episodic memories.

CLARION Cognitive Architecture

In alignment with the SOAR Cognitive Architecture's purpose, the CLARION Cognitive Architecture was proposed as a unified model for cognitive architecture. The emphases of this architecture are the interaction of implicit and explicit cognition exploration and

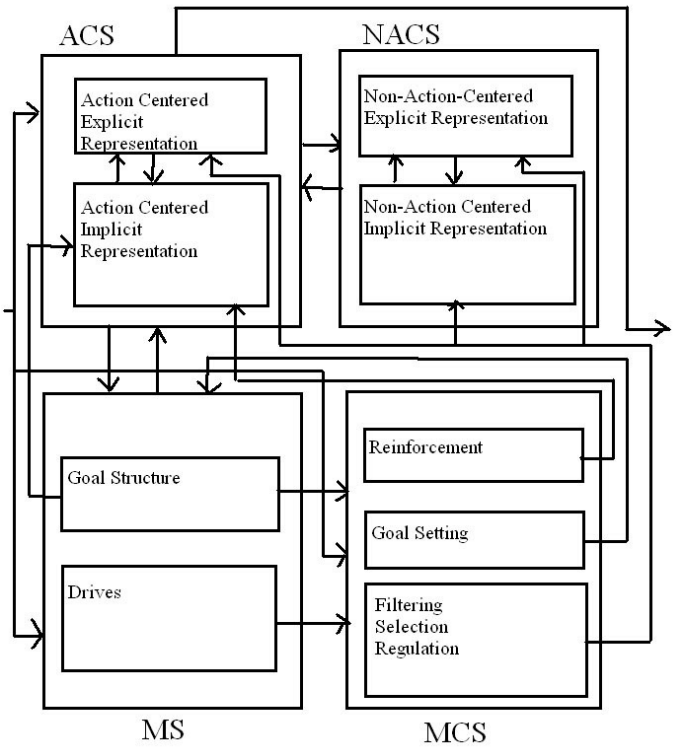


Figure 2.14: CLARION Cognitive Architecture.(Sun 2006, 2007, 2016)

bottom-up learning. CLARION Cognitive Architecture is an integrated architecture with a number of distinct subsystems with each subsystem consisting of two levels (Sun 2006, 2007, 2016) (see Figure 2.14). The top level of the subsystem encodes explicit knowledge, while the bottom level encodes implicit knowledge. All actions are controlled in the Action-Centered Subsystem (ACS), all general knowledge is represented in the Non-Action-Centered Subsystem (NACS), all goals and motivation are handled in Motivational Subsystem (MS) which affect the action of the agents and finally, the Meta-Cognitive Subsystem (MCS) acts as the system metadata. The most recent version of CLARION is CLARION 6.1.

ACT-R Cognitive Architecture

ACT-R Cognitive Architecture shares many similarities with SOAR Cognitive Architecture. It focuses on performing human cognitive tasks especially in the way human perceive, think about and act on the world (Lebiere & Anderson 2008, Anderson 1996). Similar to existing cognitive architectures, ACT-R consists of a set of components which have different functions.

The meta-process component is a complete instantiation of the system. It holds all the information regarding the time, models, version, and events. This main function of this component is to manage the whole process that occurs in the system. The model component is an abstraction level for the developer. It represents ACT-R classes that define what the classes can do as well as its own set of chunks and productions (Lebiere & Anderson 2008, Anderson 1996, Bothell n.d.). The classes include vision, motor, speech, memory and goal modules. The chunks component is the storage units of declarative memory before the data is sent and stored in the buffer component which is the interface of communication between modules in the whole system (see Figure 2.15). The most recent of ACT-R version is ACT-R 7.0 (Bothell n.d.).

2.8.2 Affect-Aware Architectures

Koko Architecture

Koko architecture (Sollenberger & Singh 2012) describes a conceptual framework to integrate affect recognition into games. The architecture has seven components connected with each other using an array of vectors (see Figure 2.16). The affect model container manages one or more affect models for each agent. The mood model manages input from the affect model and manages the agent's mood in the form of a vector. The next component is the affect repository which is the gateway through which all affective data information in the architecture passes. A similar component with the affect repository is the event repository which stores information regarding the events happened in the game environment. The sensor manager manages all the user's emotional states and environment information. This component also manages all third-party plug-ins needed by the sensors (camera, audio driver, etc.). Next, the users' information is stored and managed in the user manager. Finally, the developer interface vocabulary is a vocabulary standardisation in the system. The Koko architecture plays a role as an architectural connector, hence it is not sufficient to build a complex game as a stand-alone unit (Sollenberger & Singh 2012).

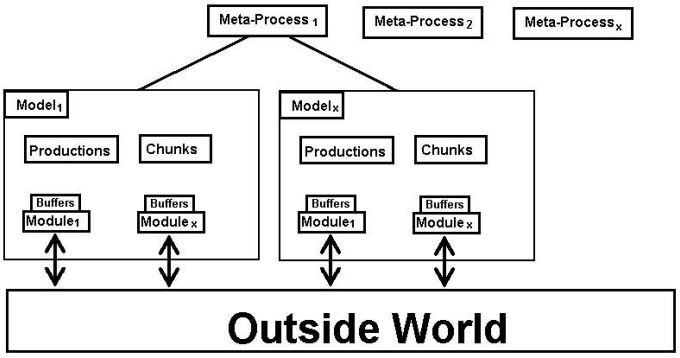


Figure 2.15: ACT-R Cognitive Components.(Bothell n.d.)

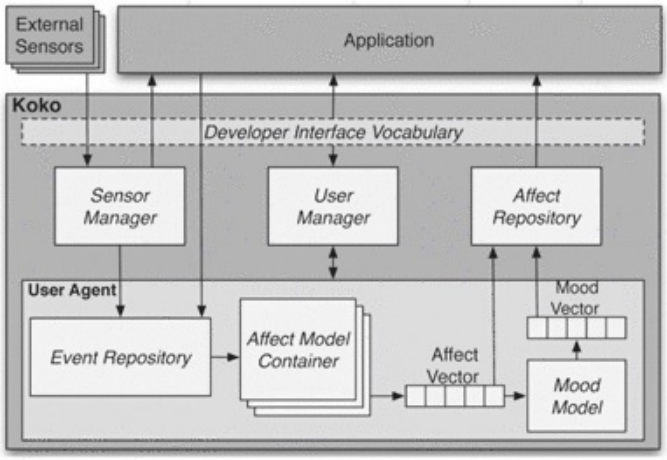


Figure 2.16: Koko Architecture.(Sollenberger & Singh 2012)

Social Signal Interpretation Framework

The SSI (Social Signal Interpretation) framework (Wagner et al. 2011, 2013) aims at the integration of multi-modal affect recognition and interpretation. The framework allows affect recognition on-line (see Figure 2.17). The framework consists of four major components. The first three components deal with sensors which are face expression recognition from a video input, voice emotion recognition from an audio input and emotion recognition from the biological signals. Each component is responsible for capturing the raw data from sensors, segmenting the data, extracting the features from the raw data and classifying them into a set of values thus detected. The framework automatically handles synchronisation between the components. Finally, the Multi-modal Fusion component deals with the fusion of the affect data from all the components. The SSI framework has been used in several projects in-

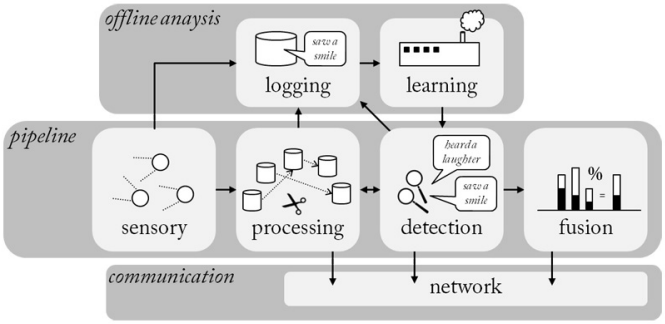


Figure 2.17: SSI Architecture.(Wagner et al. 2011, 2013)

cluding EU CALLAS Project[12], Ilhaire Project[13], OC-Trust[14], Tardis[15], ARIA-VALUSPA[16] and many more (Johannes Wagner & Jung 2009).

SAIBA Architecture

SAIBA(Situation, Agent, Intention, Behaviour, Animation) (Vilhjalmsson et al. 2016, Heylen, Kopp, Marsella, Pelachaud & Vilhjálmsson 2008, Kopp, Krenn, Marsella, Marshall, Pelachaud, Pirker, Thrisson & Vilhjlmsson 2006, Vilhjalmsson, Cantelmo, Cassell, Chafai, Kipp, Kopp, Mancini, Marsella, Marshall, Pelachaud, Ruttkay, Thrisson, Welbergen & Werf 2007) is a multi-modal behaviour generation architecture for virtual humans particularly Embodied Conversational Agents (ECAs). The architecture aims at unifying representation of ECA's communication systems. The architecture specifies the macro-scale multi-modal behaviour generation components (see Figure 2.18). Intent Planning manages the intention of the ECA's behaviour generation. Thus the intention is passed to the Behaviour Planning which manages the ECA's behavioural plan without any reference to physical behaviour. The interface between those components is called the Function Mark-up Language(FML) (Heylen, Kopp, Marsella, Pelachaud & Vilhjálmsson 2008). Next, Behaviour Realization describes multi-modal behaviours as they are realised by the final stage of the generation process. The interface between Behaviour Planning and Behavior Realization is called the Behaviour

[12]http://www.callas-new media.eu/
[13]http://www.ilhaire.eu/
[14]http://www.informatik.uni-augsburg.de/de/lehrstuehle/swt/se/projects/oc-trust/
[15]http://www.tardis-project.eu/
[16]http://aria-agent.eu/

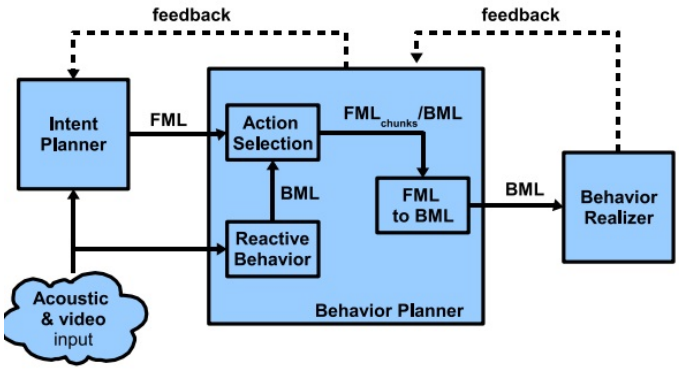


Figure 2.18: SAIBA Architecture.(Vilhjalmsson et al. 2016, Heylen, Kopp, Marsella, Pelachaud & Vilhjálmsson 2008, Kopp, Krenn, Marsella, Marshall, Pelachaud, Pirker, Thrisson & Vilhjlmsson 2006, Vilhjalmsson, Cantelmo, Cassell, Chafai, Kipp, Kopp, Mancini, Marsella, Marshall, Pelachaud, Ruttkay, Thrisson, Welbergen & Werf 2007)

Mark-up Language(BML) (Kopp, Krenn, Marsella, Marshall, Pelachaud, Pirker, Thrisson & Vilhjlmsson 2006, Vilhjalmsson, Cantelmo, Cassell, Chafai, Kipp, Kopp, Mancini, Marsella, Marshall, Pelachaud, Ruttkay, Thrisson, Welbergen & Werf 2007). SAIBA has been adopted in a number of projects including SEMAINE [17] (Schroder et al. 2012, Schröder 2010)(GRETA (Rosis et al. 2003, Bevacqua et al. 2010)), ICT VHToolKit[18] (Hartholt et al. 2013, Shapiro 2011)(SmartBody (Thiebaux et al. 2008)), ARIA-VALUSPA[19] (GRETA (Rosis et al. 2003, Bevacqua et al. 2010)) and many more.

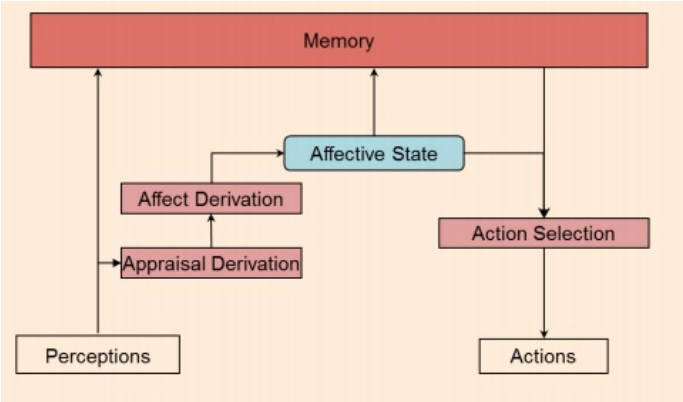


Figure 2.19: FAtiMA Architecture(Dias et al. 2011)

[17] http://www.semaine-project.eu/
[18] https://confluence.ict.usc.edu/display/VHTK/Home
[19] http://www.aria-agent.eu/

2.8.3 Other Architectures

BDI Architecture

BDI (Belief Desire Intention) Architecture is a boundedly rational agent architecture based on the BDI theory of human's rational action developed by Michael Bratman (Rao & Georgeff 1995, Wooldridge & Jennings 1995). As a boundedly rational agent, BDI focuses on the reasoning that affects the agent's actions. Furthermore, it has limited information and memory. BDI agents are required to collect information from the external environment and should need processing time due to memory limitation. BDI has three components derived from its name. Belief represents what an agent would believe about the external world. Desire are states of the agent's motivation that the agent would like to accomplish in their external environment. Finally, Intention represents what the actions are chosen based on their Desire states.

FAtiMA Modular

FAtiMA (Fearnot AffecTIve Mind Architecture) is a BDI based agent architecture designed to allow emotions and personality to influence the agent's behaviour (Dias et al. 2011). The architecture is comprised of a core layer named FAtiMA Core and the additional functional components. The core layer is a template of the agents. It works based on the emotions theory of appraisal. The data from the external world will be perceived and processed based on the appraisal theory (see Figure 2.19). In addition, a set of components are added to the architecture to handle emotion definition and generation, and the agent's states, for instance: emotion, goals, motivation, the theory of mind and culture (Dias et al. 2011).

2.8.4 Virtual Humans Tool-Kits

To boost the research in a virtual human area, the research community has been trying to create a powerful tool that contains components that can help the researchers build a virtual

human easily, so then the researchers can move to the next step of virtual human research. Building such a complicated system requires an integrated architecture consisting of shared modules or components that exhibit a range of capabilities such as those mentioned above. The SEMAINE API (Schröder 2010) and Virtual Human Toolkit (VHToolKit) (Hartholt et al. 2013) are the state-of-the-art tools consisting of integrated components to perceive, understand and exhibit behaviours. SEMAINE, an open-standards-based framework for building emotion-oriented systems, is the first integrated architecture as part of the EU-FP7 1st call STREP project. This project aimed to build a Sensitive Artificial Listener (SAL). Some years later, researchers at the University of Southern California (USC) integrated their own modules, tools and libraries to develop a virtual human together with third-party tools and released it as a toolkit under the name VHToolKit (see Table 2.2).

SEMAINE and VHToolKit are the state-of-the-art tools consisting of fully integrated components for building virtual humans. They share similarities in the components integration approach using XML-based messages. However, VHToolKit has several advantages when compared to SEMAINE in several aspects. First, VHToolKit provides a wide range of cross-platform deployment options, such as computer applications (for Windows, Linux and Mac), mobile applications (for Android and iOS) and web applications (with the flash format). Second, VHToolKit offers some API choices embedded in its framework, such as ASR APIs (PocketSphinx, Google ASR and AT&T ASR) and rendering APIs (OGRE 3D, Unity 3D, Irrlicht and Panda 3D).

Finally, VHToolKit not only focuses on the facial behaviours for both analysis and synthesis, but also body behaviours. While VHToolKit has several advantages in the input processing, output synthesis and provides excellent tools for character customisation, there is no scientific evidence on which framework has the better conceptual modelling of its virtual human. Though VHToolKit has several features that excel the other APIs for building virtual humans, there remain a number of improvements and contributions to be made including more natural expression of emotions (Bates 1994), enhanced role of personality and social relations in the selection of the agent's behaviour (Ochs et al. 2009) and better use of social signal processing (Vinciarelli et al. 2012, 2009) to drive the interaction.

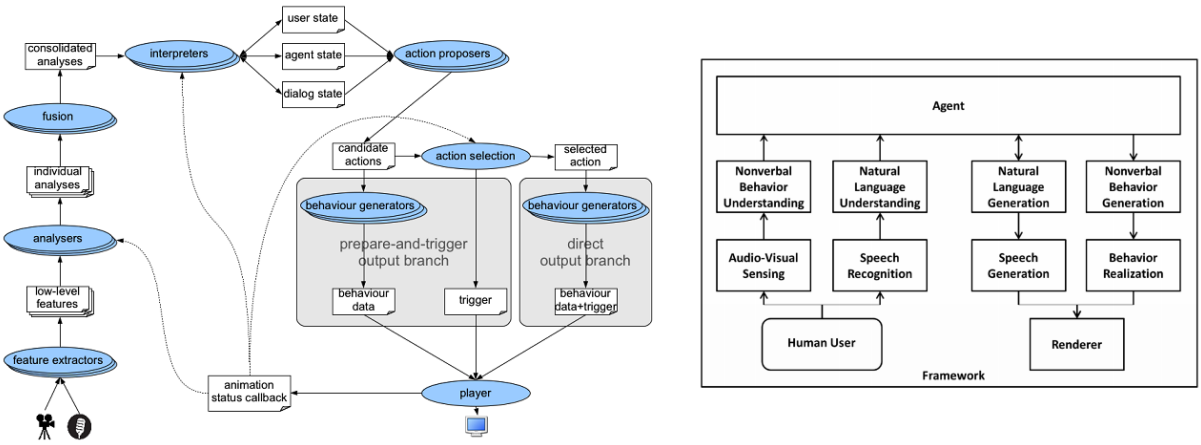


Figure 2.20: SEMAINE architecture (left) (Schröder 2010) and VHToolKit architecture (right) (Hartholt et al. 2013).

Automatic Social Signals Recognition & Analysis

Social Signals from Audio SEMAINE's audio input modules use the TUM openSMILE library (Eyben et al. 2010) for real-time audio processing and feature extraction. The user's voice is detected and captured from the microphone and is extracted for the Voice Activity Detector (VAD), prosodic analysis, the keyword spotter, non-linguistic vocalisation detector and the acoustic emotion recognition which measures valence, arousal and power of the voice. The weakness of this API is in natural language processing. The API only spots keywords from utterances. Hence it has an insufficient verbal understanding.

VHToolKit's audio input recognition and analysis are separated into two components. Non-verbal cues are processed and extracted in MultiSense components, while speech is processed and extracted in AcquireSpeech components. Its non-verbal cues recognition and analysis module in MultiSense Components uses SSI modules (Wagner et al. 2013), which are similar to SEMAINE's modules, implement the TUM openSMILE API for real-time audio processing and feature extraction. Moreover, its speech recognition module in the AcquireSpeech components connects to some compatible third-party speech recognition APIs. In contrast with SEMAINE, which has only an insufficient verbal understanding, this system excels in natural language processing by implementing third-party ASR APIs that have advantages in speech models such as PocketSphinx, Google ASR and AT&T ASR.

Table 2.2: Key properties of SEMAINE vs. VHToolKit in Automatic Social Signals Recognition & Analysis Components

Key Property	SEMAINE	VHToolKit
Non-Verbal Behaviour Analysis from Voice	Yes	Yes
Utterance Analysis	Limited	Yes
ASR API Choices	No	Yes
Non-Verbal Behaviour Analysis from Face	Yes	Yes
Non-Verbal Behaviour Analysis from Body	No	Yes
Multimodal Behaviour Analysis Fuser	Yes	Yes

Table 2.3: Key properties of SEMAINE vs. VHToolKit Synthesis Components

Key Property	SEMAINE	VHToolKit
Face Behaviour	Yes	Yes
Body Behaviour	No	Yes
Renderer API Choices	No	Yes
Multiple Characters in One Scene	No	Yes
Scene Editor	Script	Script and click based
Web Dev Support	No	Yes (flash based)
Mobile Dev Support	No	Yes (iOS, Android)

Social Signals from Video SEMAINE's video input modules acquire image data from the camera, extracts the features and then performs further analysis. They are capable of detecting nods and shakes, head tilts and emotion recognition with appearance-based Action Units (AUs) detection (Valstar et al. 2011). The appearance-based AUs detection uses 10 x 8 Uniform Local Binary Pattern histograms, which are then classified by a Support Vector Machine (SVM) into AUs. The system is only capable of detection of AU1 (raised inner eyebrow), AU2 (raised the outer eyebrow), AU4 (lowered eyebrows), AU12 (smile) and AU25 (lips parted) (Schuller et al. n.d.).

VHToolKit uses MultiSense components to track and analyse facial expressions and body posture, which is then interpreted in higher-level descriptors. In addition to the SEMAINE's video modules capabilities, MultiSense is capable of tracking and analysing user's behaviours from their body posture. MultiSense consists of several modules from SSI framework, CLM-Z FaceTracker (for multiple facial tracking with facial feature points), GAVAM HeadTracker (for 3D Head position and orientation) and FAAST API for skeleton action coding. See Table 2.2 for the automatic social signals recognition & analysis comparison overview.

Dialogue and Non-Verbal Behaviour Manager

As mentioned above, SEMAINE has an insufficient verbal understanding capability. It is intended to engage the interaction and conversation by focusing on the user's emotions and non-verbal expressions. Hence, the dialogue manager in SEMAINE concentrates on the user's non-verbal behaviours in addition to the keywords spotted in the user's dialogue. SEMAINE uses a template system and behaviour rules to manage its dialogue system. In the template system, a flexible set of easily definable templates is firstly created which specify what kind of behaviour to perform under which circumstances. A template defines a behaviour rule. The behaviour rule is determined by linking the user's previous sentences, valence and arousal from audio and the user's non-verbal behaviour from both video and audio. The rules are trained and classified using the Ripper decision tree classifier and generated using the J48 classifier.

The dialogue management in VHToolKit is handled by NPCEditor components. NPCEditor is a statistical text classifier. It is capable of responding to domain-specific conversation. In overview, the NPCEditor analyses the text of the sample utterances and responses then creates a mathematical description of the translation relationship that defines how the utterances and responses are mapped (Hartholt et al. 2013). The NPCEditor also provides a scripting ability for users to build their own dialogue managers. Moreover, the non-verbal responses are planed in NVBG components. It is a rule-based behaviour planner, similar to SEMAINE in that the rule is determined by user's non-verbal behaviours from both audio and video, in addition, the user's utterances analysis.

Social Signals Synthesis

Both APIs adopt the SAIBA standard for their multimodal behaviour generation framework. GRETA implements OGRE 3D for its renderer, while VHToolKit provides several options, for example, OGRE 3D, Unity 3D, Irrlicht and Panda 3D, resulting in the fact that VHToolKit can be deployed to the web (flash format) and mobile platforms (iOS and Android).

With regards to virtual characters rendering, SEMAINE only has one character in a scene, while VHToolKit can have several characters. Both characters in SEMAINE and VHToolKit are capable of conveying non-verbal behaviours through their facial and body expressions. They both use the same MPEG-4 standards, such as FAP for face and BAP for the body for their animation systems, however, in the SEMAINE project, the characters are only focusing on displaying face behaviours, although GRETA has the capability to synthesise body behaviours. See Table 2.3 for the comparison overview.

2.9 Virtual Humans in Game Environment

Virtual humans in games are generally represented as a manifestation of a player, called the avatar or Non-Player Characters (NPC). Virtual humans have been widely used in games as research or entertainment products. The virtual humans in games are one of the key factors to deliver the game experiences to the players along with the game aesthetic, story, and mechanics. They can be particularly interesting potential vehicles of affect as they can elicit rich emotions thus providing colourful experiences for the players. Believability in the virtual human is a virtue that helps to provide this experience. Virtual humans conveying emotions or showing behaviour at the inappropriate time would be perceived as untrustworthy characters thus affecting the interaction negatively.

Moreover, the players naturally are engaged with the NPC as part of a game. Over time, a pattern of interactions between a player and virtual human (VH) may translate into a relationship, if the game and its VH are designed to accommodate this. This thesis argues that an NPC with such capabilities will provide a new experience when playing games. Several methods have been implemented in the virtual human in the environment of the game to enhance their believability thus increasing the player experiences. Scripting is the most common approach to provide artificial intelligence to the NPC as it is the most efficient to implement and offers the flexibility of customisation (Buckland 2005). Some scripting approaches such as finite states, decision trees and scripting are commonly used to generate

NPCs behaviours. According to Ochs et al. (2009), those approaches have serious limitations when the NPC displays the same behaviours repeatedly (hence breaking the illusion of life) or when the design of the NPC's behaviours becomes increasingly more complex (to produce rich behaviours). Creating fully autonomous NPC that can react and produce their own behaviours based on their goal can be a solution to these limitations. However, it is not an easy task to build a system capable of replicating human interactive capacity.

2.9.1 Rule-Based

Rule-based systems produce outputs (e.g. behaviours) by implementing a set of rules based on an array of inputs. This is the most popular technique used to simulate human behaviours in the virtual human (Cavazza 2000, Buckland 2005, Hoekstra 2006). It is closely related to the expert systems with extensive knowledge of behaviours and the game world encoded in rules. Based on the inputs from the environments the NPC can autonomously take decisions based on the rules sets. Decision trees are one of the implementations of Rule-based techniques. They model the decisions based on the rules to weight the branches by using a tree-like graph model. The disadvantages of a Rule-based approach is that it requires a significant amount of time to craft and map all possible rules and actions. However, this problem can be solved with a machine learning approach by having a system that can automatically learn the rules from the data. Furthermore, some cognitive architectures such as SOAR (Laird et al. 1987) support the implementation of rule-based methods for the virtual agents. Example of NPCs that are implementing the rule-based method can be found in SOAR Quakebots (Laird & Duchi 2000), Civilization, Age of Empire, and Rainbow Six (Cavazza 2000).

2.9.2 Finite State Machines

Another popular technique to replicate human behaviours in the virtual human in games is the Finite State Machine. The behaviours are encoded into a set of states governed by a set

of conditions which are generally defined in a transition table. Once the conditions of a state are satisfied the current state will be updated to the next state for which the conditions are met. With this method, the states are implemented to control the virtual human behaviours based on the inputs in the current states. This method is easy to debug and quite flexible for implementation. However, human behaviour is a complex mechanism, and it is essentially impossible to design all the possible actions for the virtual human. Similarly, machine learning seems to offer solutions to this problem by learning the states and transition table from the data. There are some games that have implemented Finite State Machines for their NPCs, such as Quake, PacMan, FIFA, and World of Warcraft (Buckland 2005).

2.9.3 Machine Learning

Scripting techniques are a powerful way to mimic human behaviours in the virtual human in games. These methods, however, require very significant amounts of effort to implement. Human behaviours are complex and made up of limitless possible actions learned from past interactions. The players will be able to predict the virtual human's behaviours if the designers only provided a few rules or states of behaviours. This will negatively affect the player experiences. Yet it is seemingly impossible to craft all the possible behaviours to the virtual human. Machine learning has the opportunity to overcome those problems. It has been applied to modern commercial games for more than a decade; however, the machine learning implemented into games is still inadequate (Miikkulainen 2007). Machine learning can reduce the production costs and provide unpredictable NPCs that are able to adapt to new situations as they learn from the previous interactions. The NPCs will be capable of providing a natural way of interacting by recognising, learning and adapting from their error during interaction with the players. This will change the player's experiences in playing games (Miikkulainen 2007). Some examples of games that are implementing machine learning to their NPCs can be reviewed in (Cavazza 2000, Hoekstra 2006, Miikkulainen 2007).

2.10 Virtual Humans Interaction in the Wild

Virtual human has been successfully developed for specific scenarios and show impressive human-like capabilities when interacting with people (Kopp et al. 2005, Yuan & Chee 2003). However, the virtual human is generally designed for a specific setting and have major constraints against any scenarios outside the scenario prepared to the Virtual Humans. Hence, the virtual human applications are rarely deployed into a real-world setting as there are too many noises and an unlimited number of possibilities of the interaction between them and the players. Currently, it is not possible to craft all the possibilities of the interaction rules in the wild (i.e. real-world setting). One of the methods to create a virtual human application to interact with people in the real-world setting is to deploy a wizard-of-oz system to deal with the uncertainty in interactions. The system aims to collect a set of data (e.g. interaction rules) to be learnt by the virtual human. Moreover, by having virtual humans that able to interact with people in the wild affectively, we as researchers can learn something from the interaction.

2.10.1 Wizard-of-Oz Interaction

Wizard of Oz (WoZ) method is also commonly used for evaluating a prototype system (Green & Wei-Haas 1985*a*) as well as for initial data collection for human-computer interactions (Dahlbäck et al. 1993). Vardoulakis collected conversation data to determine the favourite topics of older adults' with the agent as well as to evaluate their system using WoZ method (Vardoulakis et al. 2012). Similarly, Sidner et al. implement WoZ method for their social agents controlled remotely by an experimenter, to collect what topics are favourable to talk about with the agent (Sidner et al. 2015). Kelley (1984) used WoZ to develop a natural language computer system as a personal assistant agent to help business professionals manage their personal calendars. Salber & Coutaz (1993) also mentioned that the WoZ method had been largely used to build large corpora for natural language processing used in retrieval information systems (e.g. reservation and advisory systems). The SEMAINE project team

created a large number of an audio-visual database of interactions with their Embodied Conversational Agent (ECA) using WoZ method, semi-automatic Sensitive Artificial Listener (SAL) and automatic SAL (McKeown et al. 2012). Concurrently, the interaction quality between the participants and the ECAs were also evaluated.

2.10.2 Interaction in a Public Space

Having interaction with a virtual human in a public space is relatively rare to be seen. Most virtual humans are deployed in a museum as a virtual tour guide to provide direction or to explain the exhibition to the visitors. Moreover, most of the interaction is done by choosing a static set of answers that are prepared beforehand. Automatic speech recognition would be extremely challenging to be implemented in a public space due to the environmental noises.

Bickmore and team designed a virtual museum guide called Tinker (Bickmore et al. 2011, 2008). Tinker is able to interact socially with the visitors. The system was installed in an open space in a human-sized body with an enormous number of interactions for two years. The participants interacted with the virtual guide with a multiple-choice touch screen. The researcher evaluated the effect of the virtual guide's social behaviour on the participants' engagement and learning in the museum. In their pilot study (Bickmore et al. 2008), 69% of visitors were completing interactions with Tinker, and 62% of them enjoyed interacting with Tinker. While in their following study (Bickmore et al. 2011), 82.3% of the 1,607 visitors indicated they were satisfied interacting with Tinker. This is due to Tinker's ability to recognise visitors faces and call the visitors by name.

Kopp and team developed Max in a static screen and installed it at a museum setting (Kopp et al. 2005, Jung & Kopp 2003). The pilot study (Jung & Kopp 2003) was done in a reasonably well-frequented hallway near to the researcher's working space before being deployed into a museum space (Kopp et al. 2005). The researcher evaluated Max by recording all the interaction between Max and the visitors in log files. The log files were then analysed for dialogues between the virtual guide and the visitors. The researchers analysed the interaction to know how many conversation and favourite topics they had between them. Similarly,

Robinson et al. (2008) also collected and analysed a corpus of favourite topics between a virtual guide and museum visitors. Finally, Swartout et al. (2010) developed human-sized twin virtual humans as museum guides, named Ada and Grace to deal with a short supply of human museum guides. They have a large but finite set of responses; Hence the visitors might get a couple of repeated answers. The Twins answered around 55.8% of questions with correct answers to the visitors.

2.11 Conclusion

Replicating human behaviour in all its complexity is not a trivial task, and thus it is unsurprising that a large number of problems are still outstanding. It is almost impossible to replicate all the human brain functions to simulate a natural and believable interaction in real time. Some intelligent VH frameworks and models have been proposed to improve the capabilities of virtual humans. The sensory systems currently available such as the SSI (Social Signal Interpretation) framework (Wagner et al. 2013), eMax (Almaev et al. 2013, 2015), and MultiSense (Scherer et al. 2012) are quite robust and can capture and process information from sensors. Moreover, the behaviour generation systems such as GRETA (Rosis et al. 2003, Bevacqua et al. 2010), and SmartBody (Thiebaux et al. 2008) are reasonably good at synthesising the embodiment of the VH. However, to process the information to perform and maintain a natural and believable interaction in real time is still considered a daunting task to researchers.

To display natural behaviour in real-time interactions, the system should accurately process the information and precisely provide an action to be performed. Inducing emotions, personality and social relationships to VH to colour the interaction is the most common method to increase the quality (e.g. naturalness and believability) of the interaction. However, to date, there are a number of theories for emotions, personality and social relationships. Most of the theories are mutually redundant, and none are designed for interoperability, leaving even more possibilities of variable combination, yet there is no comprehensive study on which

combination is the best, nor an agreed-upon framework to study this.

A number of computational models to regulate virtual human's internal states have been proposed. However, we are still missing some pieces of the puzzle to create a virtual human that provides more natural interactions in real time. In addition, they are generally not implemented to support real-time interactions, as simulations would be sufficient enough to prove the concepts of the computational models. Moreover, most of the models that implement more than one behavioural variable generally treat each variable as an independent variable, and they do not dynamically interact with each other to jointly influence the behaviour. Ideally, emotions, personality, and social relationships should interact and dynamically influence each other's values to regulate the VH's internal states and generate behaviour reflecting the states.

There are not many virtual human applications that are capable of having a natural interaction with a user deployed in a real-world setting. Interacting with a human in the wild is not a trivial task for a virtual human. There are limitless possibilities for the interactions. One of the ways to prepare the virtual human to deal with the uncertainty of the interaction in a real-world setting is to learn what are the most shown behaviours by the users during the interactions in the wild. Unfortunately, only a few studies have been done in this area (Kopp et al. 2005, Yuan & Chee 2003, Swartout et al. 2010, Jung & Kopp 2003, Bickmore et al. 2008). Hence, this thesis aims to explore the potential of the interaction between virtual humans and humans in real life. The data should potentially be used by researchers as a best practice to build virtual humans in a real-world setting.

Finally, in the area of games, there is an opportunity to improve the user experience through interactions with richer, more believable and more interesting virtual humans, usually as Non-Player Characters (NPC) in games that inhabit virtual worlds. However, there is a limitation both in terms of tools and architectures that support designing and generating such interactive behaviours for NPCs in a real-time manner. Currently, scripting of (re)action trees is commonly used to generate NPCs behaviours. Those approaches seriously break the "illusion of life" when the NPC displays the same behaviours over and over again or when

the design of the NPC's behaviours becomes increasingly more complex (in order to produce rich behaviours). Finally, current methods to build a social relationship with NPCs in games is either by a static alteration based on player's actions, game storyline or are single-event based. This limits the game experience, as it would allow the players to predict the outcome or behaviour performed by the NPC based on previous interactions.

Given the gaps mentioned, this thesis attempts to propose integrated models for personality, emotions and social relationship, derived from several theories in psychology and social sciences, for virtual human in games environment. These models allow us to adopt different approaches from the conventional behaviour scripting in the game by allowing the player to interact and forge a relationship with the NPC dynamically. To support this, several game scenarios with the models implemented are presented in this thesis. In those scenarios, the players are able to naturally build a relationship with the NPC by using their natural way for interaction (e.g. speech, facial expressions) in addition to the conventional game inputs (e.g. joystick, mouse and keyboard). With the models, the changes in their relationship are not only influenced by a traditional click-based method, but also NPC's and player's internal states, and the external environment. This thesis argues those methods will enhance the player's experience when playing a game. In addition, a data collection for behaviour rules generation is presented to design a specific game scenario.

[This page intentionally left blank]

Chapter 3

The Emotionally Realistic Social Game Agents Framework (ERiSA)

3.1 Introduction

This chapter presents the attempt to model the game companion's emotions, personality, and social relationships to enhance the game companion's believability during the interaction. A computational model is a mathematical model to simulate a phenomenon or behaviour of a complex system in the computer. We can observe the effect of the changes by adjusting the variables of the models in the simulation. Using computational models, we can model and simulate the effect of the emotions, personality, and social relationships to the VH's behaviours and apply the models in a specific scenario. Emotions proved to be an essential part of the interaction Lazzaro (2004), Hudlicka (2011), particularly in a game environment as shown in the studies described in Chapter 5 and 6. NPCs with the capability of conveying emotions to the player will induce an emotional reaction to the players leading to the enhancement of the player's game experience (Lazzaro 2004, Peters et al. 2006). Personality model contributes a unique characteristic of a game companion. This also potentially leads to the enhancement of the player's game experience. Finally, social relationships model provides the opportunity to the players to forge a social relationship with the game companions,

which conceivably leads to the enhancement of the player's immersion in the game as well as their experience during the interaction. Broekens et al. (2016) argue that one of the primary reasons why computational emotions models are rare in the virtual human, particularly in games (i.e. NPC), is usually the models adopt a particular type of Artificial Intelligence architecture as the basis for the computational models of emotions. Some models are not easy and straightforward to be implemented independently. Moreover, in this thesis, we are aiming to experiment with models, it would be difficult to do so with highly sophisticated models. We do not quite know what the best model is. Hence, we are planning to start with simple models and experimenting/analysing the models first. This has led to the definition of five requirements of the computational models.

As the requirements, firstly, the models should be relatively simple and easy to be implemented as well as to be interpretable to support an affective loop establishment between users and virtual human. This allows the virtual human to have internal states of emotions generated by multiple stimuli (e.g. their current internal states, their social relationships, their personality, user's emotions from facial expressions, etc.). Secondly, the models should also support individual uniqueness to provide a unique identity to the virtual human. This also enhances the virtual human's behaviours variation to prevent the users to predict the virtual human's behaviours. Thirdly, the models should support the evolution of social interaction over time. Fourthly, each component of the models (i.e. emotions for affective loop, personality for individual uniqueness and social relationship for the evolution of social interaction) should influence each other as literature from psychology, and social sciences suggested so (John et al. 2008, Watson & Clark 1992, Rusting & Larsen 1997, Saucier & Goldberg 1996, Funder 2001). Finally, the models should be able to calculate or quantify, and regulate the components of emotions, personality, and social relationship during the interaction between the virtual human and user.

3.2 First Version of Computational Models

The first version of Computational Models was intended to explore the potential of such models to be implemented for a virtual human. The models are kept as simple as possible, without sacrificing the purpose of the models, to make it easier to be evaluated. Personality and Emotion models are claimed to be important to maintain virtual human's believability during the interaction with human (Bates 1994, Ochs et al. 2009, Schroder et al. 2012, Li et al. 2007). In addition, social relationship model would provide the illusion that the virtual human is capable of perceiving a social relationship between them and the human during the interaction. Putting those three models together to the virtual human's behaviours, the author postulates that we can establish believable interaction between human and virtual human, by providing the ability to show a unique personality, perceive and express appropriate emotions, and show simple social skills. The proposed models of Emotions, Personality, and Social Relationship are described in subsection 3.2.1 to subsection 3.2.3.

3.2.1 Personality

Modelling personality to a virtual human is a prevailing method to enhance their believability. It is claimed as one of the key components of the virtual human's believability. As explained thoroughly in subsection 2.6.3, the OCEAN personality model is the most well-known model to be implemented into the virtual human (Li et al. 2007, Ochs et al. 2009, Schroder et al. 2012) as it describes the most details characterisation about someone's characteristics compared to others yet it is easily implemented as vectors in a computing problem. Hence, the model of personality we adopt here is based on the OCEAN (i.e. Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism) model (Saucier & Goldberg 1996). Openness is identical with the person's degree of intellectual curiosity. High openness person can be perceived as an unpredictable and open-minded person. Conscientiousness reflects the person's level of organisation and dependency. High conscientiousness people can be perceived as stubborn and organised. Extraversion indicates the level of positive

energy and emotions of a person. An extrovert person is generally perceived as an energetic and positive person. Agreeableness implies the degree of cooperativeness of a person. High agreeableness person is submissive. Finally, Neuroticism indicates the level of negative energy and emotions of a person. High Neuroticism person is typically perceived as an emotionally unstable person and vice versa. The values of the traits are represented by a set of real numbers between -1 to 1, where 1 represents the strongest possible value in that particular trait. For example: if the agent has a personality with an extraversion value of 0.65 and conscientiousness of -0.65 then the agent tends to be talkative and shows a lack of self-discipline. Our model is represented as:

$$\mathcal{P} = \{O, C, E, A, N\} \in [-1, 1]^5 \tag{3.1}$$

3.2.2 Social Relationship

As mentioned in subsection 2.6.5, a social relationship can be posed into a dimensional computational problem. There are several dimensions proposed to model virtual human's social skills. In this thesis, the author is interested in two specific dimensions of the social relationship. The one that explains how close one person to another is, the other one describes the social attitude of a person toward another. Hence, in this social relationship model, two variables of social relationship are highlighted: *Affinity* represented with $\mathcal{A}$ and *Familiarity* denoted with $\mathcal{F}$. *Affinity* denotes the degree of liking between the player and NPC at time t. The change in the degree of *Affinity* ($\mathcal{A}_t \in [-1, 1]$) at time t depends on the emotion v_t of the game companion towards the player at that time (explained below), so a negative emotion at time t will lead to the decrease of *Affinity*. The value of Affinity also is driven by the level of significance the event has to the social relationships between the game companion and player, denoted by $\mathcal{S}$. It is to indicate how important this event to the relationship between them. The more important the event, the higher the impact for the game companion's *Affinity* towards the player. The degree of *Familiarity* depends on how many times the game companion and the player met, which in our game scenario

is measured in terms of how often they play the game. *Familiarity* at time t indicates how intimate the player is to the NPC, given by the number of interactions and time spent between two individuals. The value of *Familiarity* is normalised to a real number between 0 to 1 ($\mathcal{F}_t \in [0, 1]$). The value is initialised with 0 assuming the NPC and player never met before ($\mathcal{F}_t = 0$ leads to $\mathcal{R}_t = 0$) and will eventually reached 1 (see Equation 3.11). $\mathcal{F}$ normally is used as a measurement of how well someone knows each other. However, to simplify, the value of $\mathcal{F}$ in this thesis is defined by the frequency of two people have met. The direction of the Social Relationship depends on the value of *Affinity* ($\mathcal{A}_t \in [-1, 1]$) of the NPC towards a player at time t. If the value of $\mathcal{A}_t \geq 0$, the relationship between them going towards positive direction, while if the value of $\mathcal{A}_t < 0$ indicates they are having a negative relationship between them (e.g. hatred). The value of the Social Relationship at time t ($\mathcal{R}_t \in [-1, 1]$) is calculated as:

$$\mathcal{R}_t = (\mathcal{A}_t \times \mathcal{F}_t) \tag{3.2}$$

where

$$\mathcal{A}_{t+1} = \mathcal{A}_t + \Delta\mathcal{A} = \mathcal{A}_t + v_{t-1} \times \mathcal{S} \tag{3.3}$$

3.2.3 Emotion

Both personality and emotion play an important role in the individual's behaviour (Hilgard 1953, Picard 1997). For example, an individual with high extraversion tends to express and perceive positive emotions, while an individual with high neuroticism trait tends to express and perceive negative emotions. In addition to personality, the social relationship also plays a role in the individual's emotional behaviour. To model emotion, we adopt the Dimensional Affect theory of Russell (Russell 1980). For simplicity, we only use valence, which indicates how positive or negative a person is feeling. Subsequently, the game companion's emotion at time t ($v_t \in [-1, 1]$) is computed based on the average of events $\mathcal{E}_i$ (e.g. per-

ceived player's emotions from their facial expressions, player's utterances, player's choices of words/sentences) at that time in addition to personality ($\mathcal{P}_E$ for extraversion and $\mathcal{P}_N$ for neuroticism) and social relationship at time t ($\mathcal{R}_t$). To model the dynamic properties of emotion perception in a simple yet convincing way, we implemented a decay function to our model. In the absence of any emotion-evoking stimuli, the value of an affective dimension is assumed to decay over time (Picard 1997). The value of the decay rate $r \in [-1, 1]$ is set based on the game companion's personality. In the absence of any internal mechanisms of the agent such as goals and resources, we determine the emotion of the agent to depend on the User entirely

$$v_t = max(-1, min(1, v_{t-1} \times e^{-rt} + \mathcal{I}_t)) \tag{3.4}$$

where intensity of emotions at time t ($\mathcal{I}_t \in [-1, 1]$) is represented by:

$$\mathcal{I}_t = \frac{\sum_{i=1}^{n} \mathcal{E}_{i,t}}{n} \times (\mathcal{P}_E - \mathcal{P}_N + \mathcal{R}_t) \tag{3.5}$$

3.3 Second Version of Computational Models

The first version of computational models of emotions, personality, and social relationships is a simple version of the current computational models as it was intended to explore the initial study and data collection for the game companions. The models were implemented in the first version of The Smile Game, described in Section 4.5. However, it did not clearly show how the game companion's internal states such as emotions, personality, and social relationships influence the gameplay. Based on the findings from the pre-preliminary study presented in Section 4.5 and 4.6, we propose the second version of computational models of emotions, personality, and social relationships. The representation of the personality model is identical with the first one. While the enhanced models of emotions, social relationship, as well as the interaction between emotions, personality and, social relationships, are thoroughly

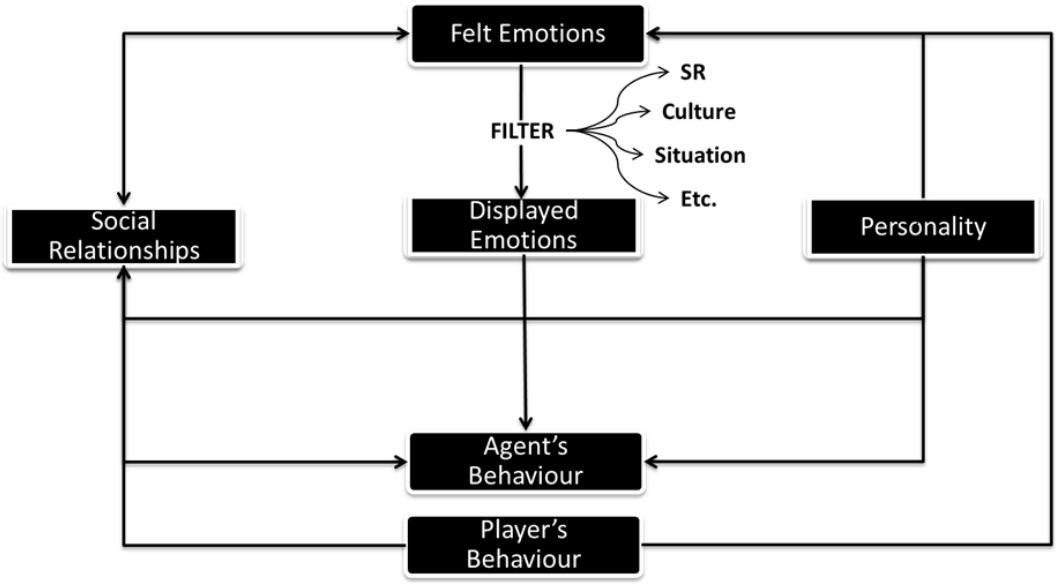


Figure 3.1: Models of Personality, Emotions and Social Relationship

explained in this section.

Figure 3.1 illustrates how emotions, personality, and social relationships interact with each other on influencing the game companion's behaviour in our models. We distinguish internal emotions that the game companion is actually "felt" from what they displayed. Some aspects, such as their social relationship with the user, cultural display rules, and the current situation can influence people to mask their emotions (Ekman 2007, Jack et al. 2012, Shattuck 1988). The interaction and influence between components as well as the underlying concepts behind the models are presented in subsection 3.3.5.

3.3.1 Model of Emotions

There are several approaches to emotion classification in psychology. Classification helps distinguish one emotion from another. We based our emotion model on the most classic and common emotion classification system: the six universal emotions expressions, proposed by Paul Ekman (Ekman 1999*b*,*a*, Ekman & Friesen 1969). He reinvestigated Darwin's work and claimed that there are six universal emotions (later extended to 7, adding contempt), which are produced and recognised independently of cultural background. The emotions are anger, disgust, fear, happiness, sadness, and surprise and they have evolved in the same way for all

humankind. Moreover, the theory suggests that our facial expressions show these emotions and the display of emotions are produced and recognised universally, i.e. independent of culture or other nurture effects (Valstar 2008). The model can be simplified to positive and negative emotions where anger, disgust, fear, and sadness can be classified into negative emotions, happiness can be categorised as a positive emotion, and lastly, surprise can be either positive or negative emotion depending on the trigger or event. The intensity of emotions in our model is described in Equation (3.6).

$$v = \begin{cases} \{anger, disgust, fear, sadness\} \in [-1, 0]^4 \\ \{happiness\} \in [0, 1] \\ \{surprise\} \in [-1, 1] \end{cases} \tag{3.6}$$

3.3.2 Emotions Threshold, Intensity, and Saturation

The threshold for displaying emotion and the emotion saturation (i.e. maximum level) are vary depending on personality. In our work, the value of saturation for all emotions is set to -1 for negative emotions and 1 for positive emotions, while the value of emotion threshold depends on the game companion's personality. Exceeding the threshold triggers the activation of the game companion's emotion display, while the saturation level defines the maximum value of the agent emotions variable. A rapid repeated activation causes the intensity of the emotion to rise to the maximum of its saturation level. For example, an NPC with a threshold of 0.6 and saturation of 1.0 for happiness, will feel happy and change their behaviour gradually when intensity for happiness exceeds the value of 0.6 until the value of saturation is reached. In our models, personality sets the basic value of the emotion threshold (θ_v). In addition, the emotions' threshold at time t is dynamically changed depending on the social relationship between i and j. Equation 3.7 represents the model of emotion threshold $\mathcal{T}_{v,i,j,t}$.

$$\mathcal{T}_{v,i,j,t} = \theta_v \times \mathcal{R}_{i,j,t-1} \tag{3.7}$$

A decay function is applied in our work, allowing emotion to gradually return to the rest state over a period of time. The emotion intensity will fall below the level of the threshold, thus disappearing unless it is re-activated (Picard 1997, Velsquez 1997). To give an illustration of emotion decay, imagine that the activation of emotion is a rock that hits a pool. The ripple will gradually vanish over a period of time unless another rock hits the pool once more. Moreover, repetitive strikes of rocks in the same place (equivalently to striking with the same emotion) will have the effect of escalating the intensity of the ripple. Similarly, to the ripple, repetitive activation of emotion will have the effect of increasing the intensity of the emotion. We based our decay function on a sigmoid function proposed in Steunebrink et al. (2008), with $r \in [-1, 1]$ as the decay rate, $\mu = ln(2)/r$. Intensity of emotion v at time t (see Equation 3.8 and 3.9) depends on the emotion triggered by the game companion as result of the event $\mathcal{E}_{v,t}$, the importance of the event (Frijda 1988) ($\mathcal{M}_{v,t}$), the personality $\mathcal{P}_p$ of the game companion and how strong the personality influences the intensity of emotions ($\mathcal{W}_p$) at time t.

The emotion models proposed, however, poses a problem when the designers or developers design a series events repetitively with a considerable number for every event. The value of the intensity of the emotion will saturate quickly resulting to the same event will have no effect anymore to the virtual human, until it decays overtime or provoked by events that have opposite effect to the virtual human's current internal state of emotions.

$$\mathcal{I}_{v,i,j,t+1} = \begin{cases} min(1, max(0, \mathcal{I}_{v,i,j,t+1} + \mathcal{I}_{v,i,j,t})) & v \geq 0 \\ max(-1, min(0, \mathcal{I}_{v,i,j,t+1} + \mathcal{I}_{v,i,j,t})) & v < 0 \end{cases} \tag{3.8}$$

where

$$\mathcal{I}_{v,i,j,t} = \begin{cases} \frac{\sum_v^n \mathcal{E}_{v,t} \times \mathcal{M}_{v,t} + (\mathcal{P}_E \times \mathcal{W}_p)}{1 + e^{(t-\mu)r}} - \mathcal{T}_{v,i,j,t} & \mathcal{E} \geq 0 \\ \frac{\sum_v^n \mathcal{E}_{v,t} \times \mathcal{M}_{v,t} - (\mathcal{P}_N \times \mathcal{W}_p)}{1 + e^{(t-\mu)r}} - \mathcal{T}_{v,i,j,t} & \mathcal{E} < 0 \end{cases} \tag{3.9}$$

3.3.3 Model of Social Relationships

Similar with the first models, we adopt the two dimensional social relationship model for NPCs in games based on (Ortony 1991, de Rivera & Grinkis 1986, Altman & Taylor 1973, Hays 1989) with the *Familiarity* in the x-axis and *Affinity* in the y-axis. The model dynamically interacts with the game states and rules, creating dynamic gameplay based upon the social relationship together with the game companion's personality and emotional states. *Affinity* at time t denotes the degree of liking between the person i and j ($\mathcal{A}_{i,j,t}$), while *Familiarity* at time t indicates how intimate the person i is to person j ($\mathcal{F}_{i,j,t}$), given by the number of interactions and time spent between two individuals.

3.3.4 Familiarity and Affinity

Equation 3.10 describes our model of social relationship. The value of *Familiarity* in our social relationship model is governed by the frequency of the interactions characterised by time spent together by two entities. Research shows the frequency of interactions has a strong correlation with the social relationship strength (Hays 1989, Sutcliffe & Wang 2012, Altman & Taylor 1973). The more frequent the interactions, the stronger the social relationship between two people. The values is calculated based on the number of interaction between person i and j at time t ($\mathcal{X}_{i,j,t}$) divided by the *Familiarity* saturation of person i towards person j, denoted by $\mathcal{Y}_{i,j}$. The increasing of the value of *Familiarity* saturation technically would make it harder for people to forge a social relationship with the agent (i.e. take a longer time to build). Moreover, we could theoretically vary this value for different agents. However, in this thesis, we keep it fixed for the initial observation. Studying the effect of agents having different values for *Familiarity* saturation could be interesting future work. The value of *Affinity* in our model represents the attitude of person i towards person j at a particular time t. This value is driven by the emotion intensity perceived by the game companion during the event $\mathcal{E}_{v,t}$ and level of significance this event has to the social relationships between person i and j ($\mathcal{S}_{i,j}$), aggregated to the *Affinity* of time $t-1$. The model translates negatives emotions (i.e. anger, disgust, fear, sadness, and negative surprise)

will lead to a negative *Affinity* (or will diminish the value of a positive *Affinity*). On the other hand, positive emotions (i.e. happiness, and positive surprise) will lead to an increase of a positive *Affinity* of a person i towards a person j. Similarly to the emotion model proposed, a decay function is applied to this model as well, where $r \in [-1, 1]$ as the decay rate and $\mu = ln(2)/r$ as the half-life. A social relationship takes lots of hard work to maintain. All relationships require persistent interactions for maintenance. A relationship without further interaction will gradually decrease in a period of time, while multiple positive interactions lead to a positive direction for social relationships between 2 people and vice versa. Finally, In our models, we introduce a variable Social Relationships Intention from person i to person j at time t, denoted by $\mathcal{N}_{i,j,t}$. The Social Relationships Intention variable is inspired by the motivation of person i to build social relationship towards person j. The variable allows each individual to have different motivation or intention level to build a social relationship with someone depending on their own goals (Maslow 1943, Hennig-Thurau et al. 2002, Dweck & Leggett 1988).

$$\mathcal{R}_{i,j,t} = \frac{\mathcal{F}_{i,j,t} \times \mathcal{A}_{i,j,t} \times \mathcal{N}_{i,j,t}}{1 + e^{(t-\mu)r}} \in [-1, 1] \tag{3.10}$$

where

$$\mathcal{F}_{i,j,t} = \frac{\mathcal{X}_{i,j,t}}{\mathcal{Y}_{i,j}} \in [0, 1] \tag{3.11}$$

and

$$\mathcal{A}_{i,j,t} = \mathcal{A}_{i,j,t-1} + (\mathcal{S}_{i,j} \times \mathcal{I}_{v,i,j,t-1}) \in [-1, 1] \tag{3.12}$$

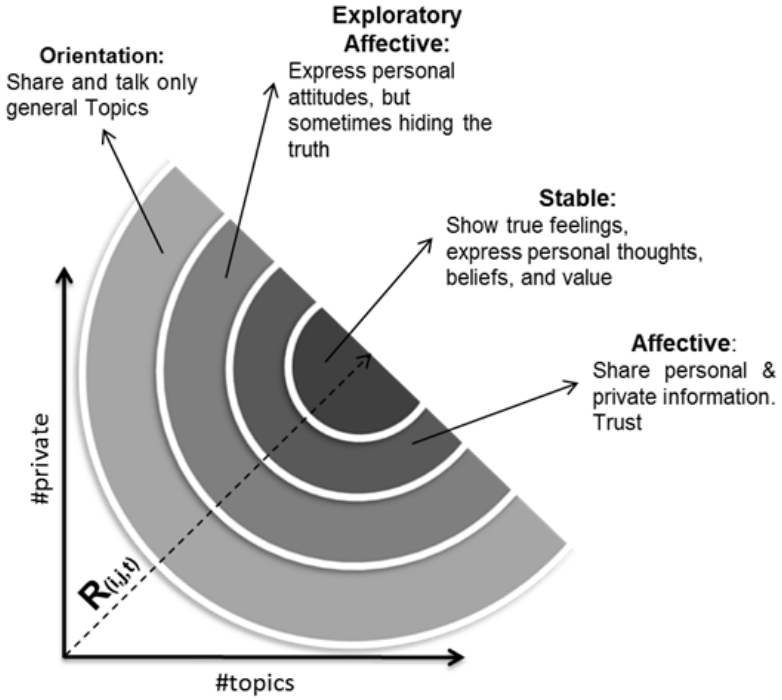


Figure 3.2: Impacts on Behaviours from Social Relationships Model

3.3.5 Interactions between Component

Personality Effects on Felt Emotions

Based on John et al. (2008), Watson & Clark (1992), Rusting & Larsen (1997), personality has a certain role to alter emotions' intensity. In this work we only focus on two components of personality: Extraversion and Neuroticism. High extraversion person tends to express and appraise positive emotions, while an individual with high Neuroticism tends to express and appraise negative emotions (John et al. 2008). Hence, the intensity of positive emotions of person i towards j at time t ($\mathcal{I}_{v+,i,j,t} \in [0,1]$) is multiplied for high Extroversion NPCs, while it has no effect for high Neuroticism NPCs. On the other hand, the intensity of negative emotions of person i towards j at time t ($\mathcal{I}_{v-,i,j,t} \in [-1,0]$) is multiplied for Neuroticism NPCs, and has no effect for high Extroversion NPCs. Moreover, in our framework, personality has the role of determining the emotion activation threshold. It impacts the game companion's emotional reaction to the same event with differing intensities (Frijda 1988). Game companion with high Neuroticism have a higher threshold for positive emotion activation ($\theta_{v+} \in [0,1]$) and lower threshold for negative emotion activation ($\theta_{v-} \in [-1,0]$)

compared to the ones with high Extroversion personality. The details description of the models of the emotions threshold and intensity is presented in Section 3.3.2.

Personality Effects on Social Relationships

Personality plays some part in determining the goal in relationships (John et al. 2008). A person with a positive Extraversion trait tends to be sociable, friendly and has many friends. Similarly, a person with a reverse Neuroticism trait ($\mathcal{P}_{\mathcal{N}} < 0$) is committed to greater relationships. On the contrary, a person with negative Extraversion or positive Neuroticism is someone who is reserved, likes to be alone and has few friends. Section 3.3.4 illustrates the impact of $\mathcal{N}_{i,j,t}$ on relationships between person i to person j at time t.

Personality Effects on Behaviour

Personality becomes one of the aspects that influences a virtual human's behaviour. Personality characterises the behavioural patterns of an individual (Saucier & Goldberg 1996, Funder 2001). In our models, we used the Big Five Inventory (John et al. 2008) as a pattern on how agents would behave and react to the environment. For example, their choices of words and sentences. Extraversion person tends to be active, talkative, energetic, noisy, adventurous and sociable. While, a person who has reversed Extraversion trait is a quiet, reserved, shy, withdrawn and silent character. Table 3.1 shows certain characteristic of particular traits.

Emotions Effect on Social Relationships

Emotions play role on relationships between people. Positive emotions induce an increase to the degree of *Affinity* a person has for another, while negative emotions will lead to the decrease of the degree of *Affinity*. In our model, the intensity of felt emotions towards an event $\mathcal{E}_{v,t}$ at a time t ($\mathcal{I}_{v,i,j,t+1}$) modifies the degree of *Affinity* a person i has towards a

person j (see Equation 3.10), hence negative emotions at time t will lead to the decrease of the degree of *Affinity* of person i to j ($\mathcal{A}_{i,j,t}$).

Table 3.1: Personality Characteristic

Extraversion ($\mathcal{P}_\mathcal{E}$)		**Neuroticism ($\mathcal{P}_\mathcal{N}$)**	
Pos.	**Neg. (R)**	**Pos.**	**Neg. (R)**
Active	Quite	Tense	Stable
Talkative	Silent	Anxious	Stable
Energetic	Shy	Nervous	Calm
Noisy	Withdrawn	Moody	Contented
Adventurous	Retiring	Temperamental	
Sociable	Reserved	Emotional	

Social Relationships Effect on Behaviour

Figure 3.2 exhibits the impact of social relationships on behaviour shown by a game companion in our models, based on the work of Altman and Taylor (Altman & Taylor 1973). In our work, the social relationship level between person i and j at time t defines the breadth and the depth of the topics during the conversations between the individuals. The first stage is an Orientation phase, where each individual starts to build a relationship by engaging in small talk. Then, in the stage of Exploratory Affective, each individual starts to express weak personal attitudes regarding a topic. In this stage, people generally start to open up on moderate topics, but they do not express the whole truth yet. In the Affective phase, people start to express their real feeling. In addition, they usually start discussing their personal and private lives and safely begin to argue with each other without necessarily negatively affecting their relationship. Finally, they will reach the final stage: A Stable stage, where the individuals express personal thoughts, beliefs, and values among them (Altman & Taylor 1973). Hence, the higher the value of the social relationship between person i to j at time t ($\mathcal{R}_{i,j,t}$), the broader and the more profound the topics of the conversations will be between those persons.

3.3.6 Overview of the Variables

Table 3.2 lists all the important variables of the computational models proposed. There are eighteen main variables which presented throughout this chapter complete with the possible value for the variables. The list organised as follow: the variables used in the Emotions computational models followed by the Personality and Social Relationship models, variables to describe Event and followed by the global variables that applied to all the computational models (e.g. Euler constant and decay rate).

Table 3.2: List of Variables

Variable	**Details**	**Values**
v	NPC's valance of emotions	$\mathbb{R} \in [-1, 1]$
$\mathcal{I}$	NPC's intensity of emotions	$\mathbb{R} \in [-1, 1]$
$\mathcal{T}$	NPC's dynamic emotion threshold	$\mathbb{R} \in [-1, 1]$
θ	NPC's basic emotion threshold	$\mathbb{R} \in [-1, 1]$
$\mathcal{P}$	NPC's Personality	$\{O, C, E, A, N\} \in [-1, 1]^5$
$\mathcal{W}$	Personality weight	$\mathbb{R} \in 0, 1]$
$\mathcal{R}$	Social Relationship (SR) between NPC and Player	$\mathbb{R} \in [-1, 1]$
$\mathcal{A}$	NPC's Affinity towards a player	$\mathbb{R} \in [-1, 1]$
$\mathcal{F}$	NPC's Familiarity towards a player	$\mathbb{R} \in [0, 1]$
$\mathcal{N}$	NPC's intention for SR towards a player	$\mathbb{R} \in [-1, 1]$
$\mathcal{X}$	Number of interaction between NPC and a player	$\mathbb{Z}^+$
$\mathcal{Y}$	Familiarity saturation of a player with NPC	$\mathbb{Z}^+$
$\mathcal{S}$	Event's significance to SR between NPC and Player	$\mathbb{R} \in [0, 1]$
$\mathcal{M}$	Event's significance to the NPC's emotion	$\mathbb{R} \in [0, 1]$
$\mathcal{E}$	Event's value	$\mathbb{R} \in [-1, 1]$
e	Euler number	2.718...
r	Decay rate	$\mathbb{R} \in [-1, 1]$
μ	Half life rate	$\mathbb{R} \in [-1, 1]$

3.4 Simulation of The Models

Simulations with two scenarios were created to evaluate the models before implementing them into the system. The first scenario was a situation where a player is playing The Smile Game (see Chapter 5) with two characters who represent both extreme side of personalities. While the second scenario was a sequence of events, where a player was interacting with those two characters.

3.4.1 First Scenario: Play The Smile Game with Poppy and Spike

Before evaluating the system with real players, we created a simulation of a game with the first version of computational models. We experimented with two SEMAINE characters: Poppy and Spike (McKeown et al. 2012) to simulate a positive emotion (i.e. happiness) leading to the activation of smile behaviour. We modelled Poppy to have high Extraversion ($\mathcal{P}_{E,N} = \{0.5, 0.2\}$) personality, while Spike was modelled to have Neurotic person trait ($\mathcal{P}_{E,N} = \{0.2, 0.5\}$). In addition, the relationship value $\mathcal{R}$ between the simulated player and the game companions was set to 0.65 and decay rate r to 0.02 for both game companions. We performed random positive interactions with both characters for a specific time. Both characters were receiving the same sequence of interactions.

Figure 3.3 shows the result of the simulation of both characters positive emotions ($\mathcal{I}_t$) as a result of three positive events (e.g. received a favourite item, perceived a smile from player's facial expressions, greetings or hearing jokes). The first event was a series of medium intensity values of positive interactions from the player from $t = 13$ to 17 seconds, the second one was a series of high-intensity values of positive interactions from the player from $t = 25$ to 31 seconds and the last event was a series of low-intensity values of positive interactions from the player from $t = 41$ to 60 seconds.

The θ_{S1} line indicates the threshold of generating a small smile by the game companions, while the θ_{S2} line indicates the threshold for an uncontrollable smile. In this simulation, we used the same value of both thresholds for both characters. As we can see in the figure 3.3, from $t = 6$ to 11 seconds, Poppy who has a high extraversion value showed a small smile and in the $t = 26$ seconds, Poppy was unable to hold her smile at the end, while Spike who has a high neuroticism only just showed a small smile at the second interactions. The simulation also showed the emotion's intensity was decreasing when there was no interaction performed between them. The simulation of the models is significant as a pilot study of the first game scenario in this thesis described thoroughly in the Section 4.5.

Figure 3.4 illustrates the result of the simulation of a short interaction with both characters.

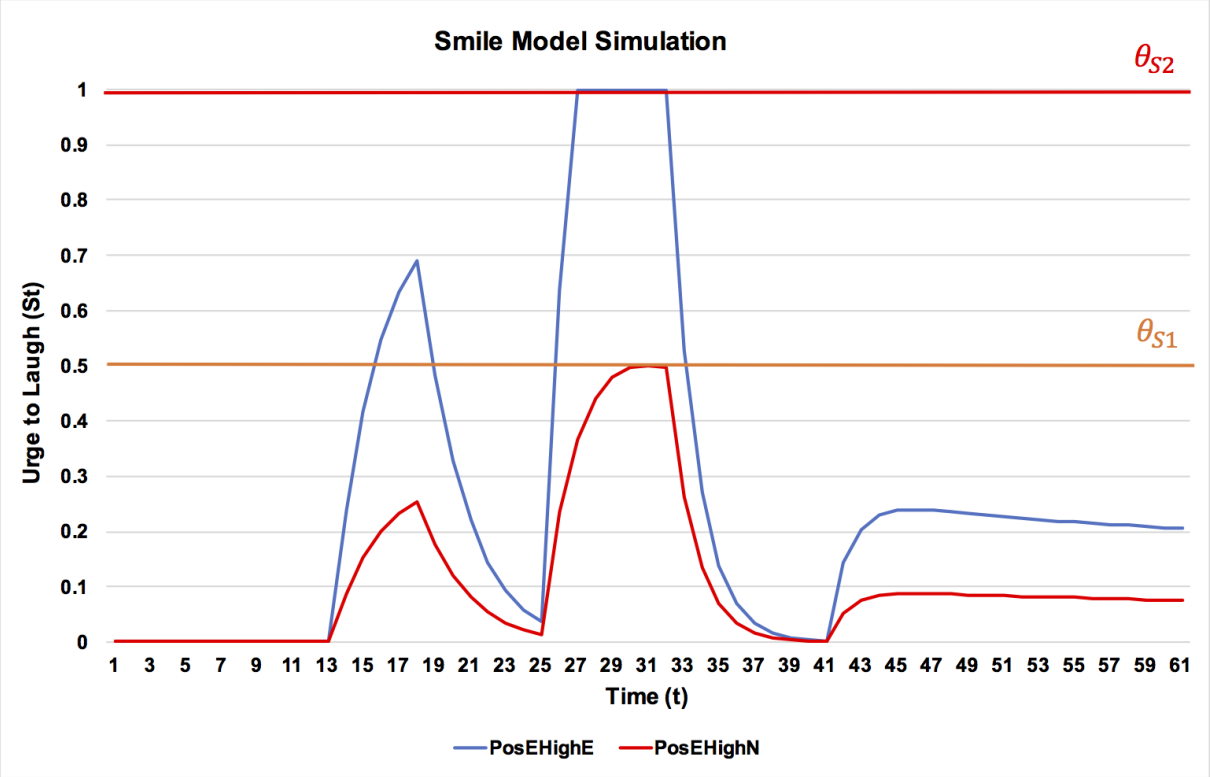


Figure 3.3: Smile Simulation. θ_{S1} line indicates the threshold of generating a small smile by the game companions and θ_{S2} line indicates the threshold for an uncontrollable smile.

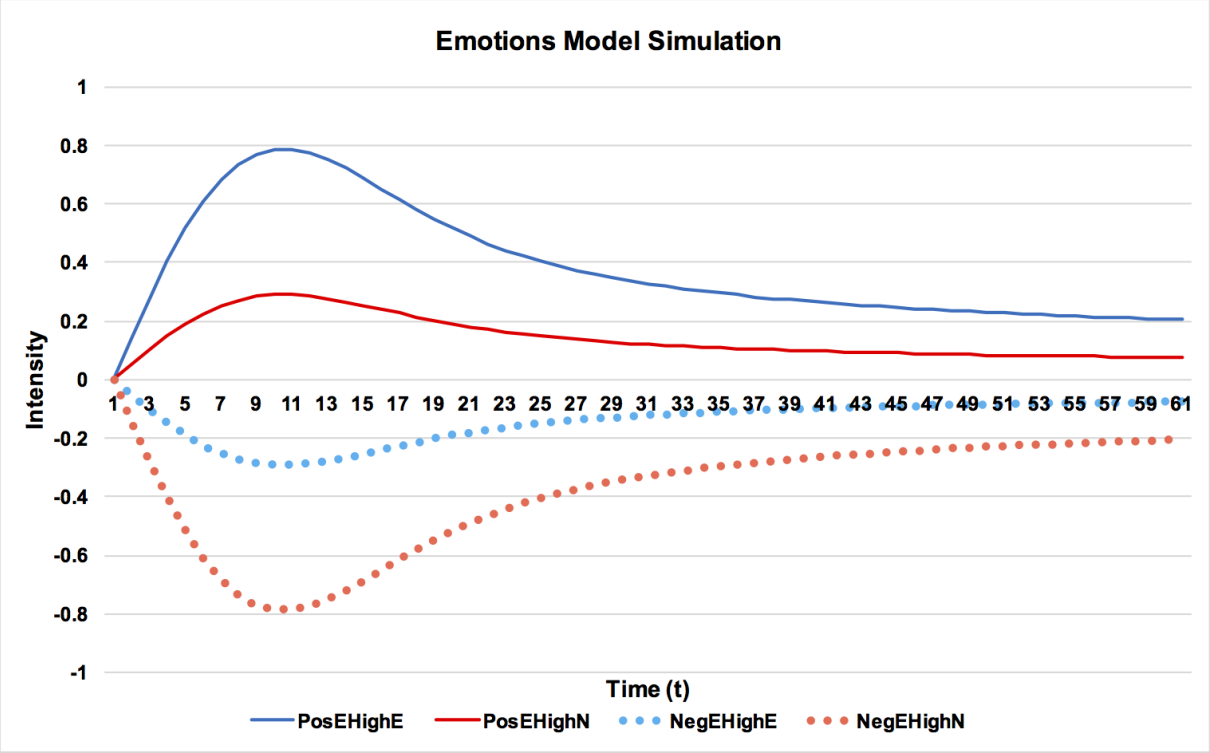


Figure 3.4: Distribution of Emotions Models - Short Interaction

Solid lines indicate positive interactions with the characters while dotted lines represent negative interactions with both characters (blue for Poppy and red for Spike). The first event occurred from $t = 0$ to 11 seconds, and the second event illustrates an absent of the interaction from $t = 12$ to 60 seconds. As shown in the figure, Poppy has a greater area of positive intensity than Spike in the positive interactions, while Spike has a greater area of negative intensity than Poppy in the negative interactions.

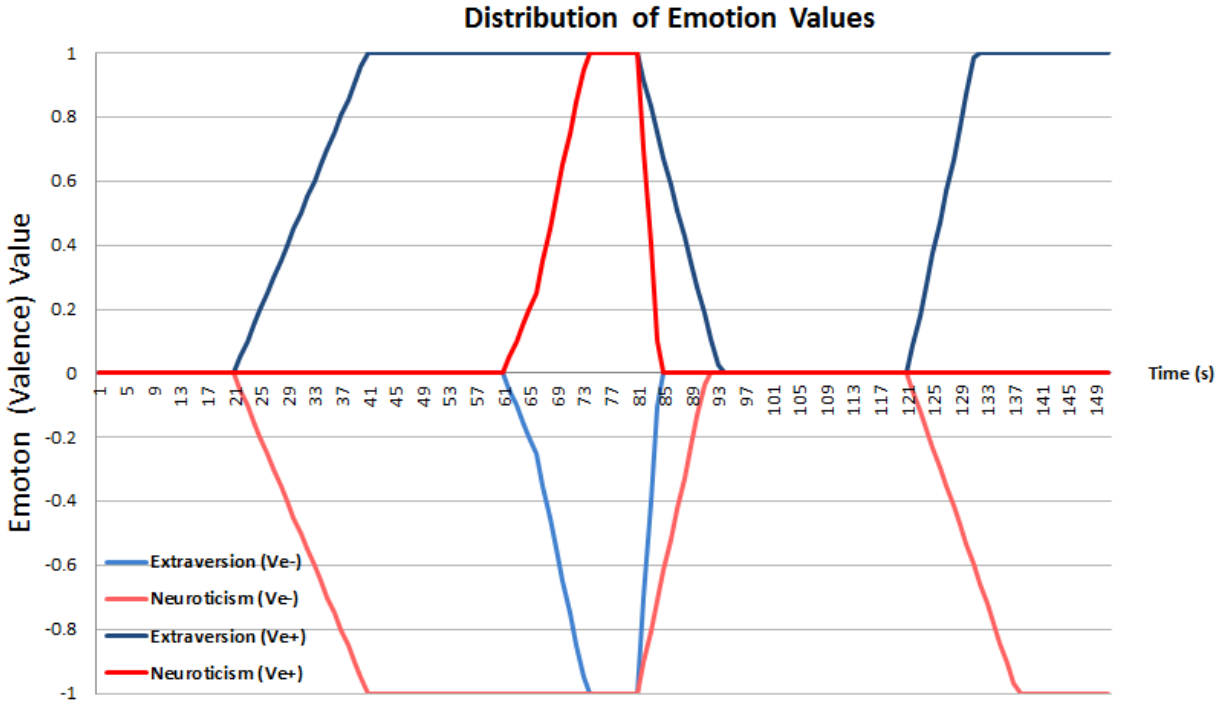


Figure 3.5: Distribution of Emotions Models - Long Interaction

3.4.2 Second Scenario: Interaction with Poppy and Spike

Two simulation experiments were performed to demonstrate the effectiveness of the models for interaction with the players with the second version of computational models. In the first simulation experiment, we simulated two events perceived by the game companion simultaneously at a particular time t. To simulate positive interactions, we set the values of $\mathcal{E}_{v+,t}$ from the range of 0.1 to 0.5 with the importance level of the event to the game companion, $\mathcal{M}_{v,t}$ ranged from 0.1 to 0.4, between time $t = 1$ to $t = 80$ and 101 to 150. Furthermore, to simulate negative interaction, a set of values $\mathcal{E}_{v-,t}$ between -0.1 to -0.5 with the same range of values for the importance $\mathcal{M}_{v,t}$ and time set to the positive interactions,

are applied. In the second experiment, we simulated two positive events perceived by the game companions simultaneously at a particular time t for social relationship values. All the events value $\mathcal{E}_{v+,t}$ and the importance level of the event to the game companion, $\mathcal{M}_{v,t}$ are set to a static value of 0.5 with an absence of the interactions between time $t = 780$ to $t = 950$.

The simulation parameters for both experiments are: the number of saturation of familiarity $\mathcal{Y}_{i,j,t} = 100$, $\mathcal{N}_{i,j,t} = 0.6$, decay rate for social relationships models $r = -0.1$, $\mathcal{M}_{\sqsubseteq,\sqcup} = 0.02$, $\mu = ln(2)/r$, emotion decay rate $r = -0.5$, negative emotions' basic threshold for high Extraversion $\theta_{v-} = -0.3$, positive emotions' basic threshold for high Extraversion $\theta_{v+} = 0.2$, negative emotions' basic threshold for high Neuroticism $\theta_{v+} = -0.2$, positive emotions' basic threshold for high Neuroticism $\theta_{v+} = 0.3$, $\mathcal{W}_p = 0.2$. The simulation also showed the emotion intensity was decreasing (or increasing for negative emotions) when there was no interaction performed between them.

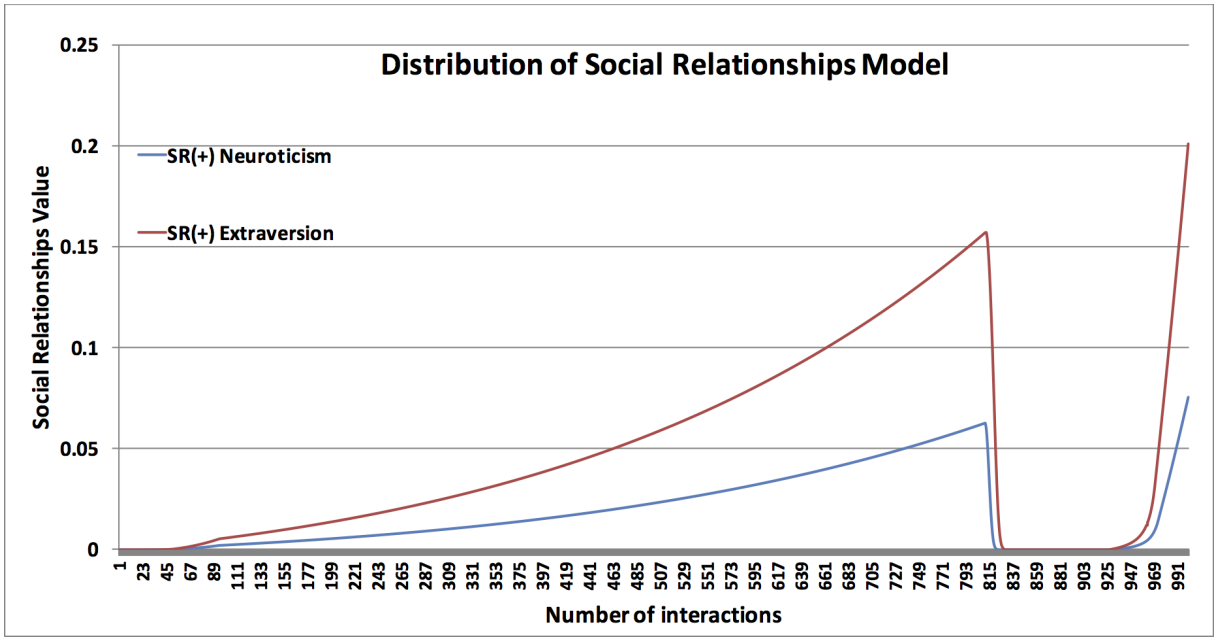


Figure 3.6: Distribution of Social Relationships Models

Figure 3.5 shows the distribution of emotions intensity values $\mathcal{I}_{v,i,j,t+1}$ on the influence of positive and negative emotions on both high Extraversion $\mathcal{P}_{E,N} = \{0.5, 0\}$ and high Neuroticism $\mathcal{P}_{E,N} = \{0, 0.5\}$ game companions for 150 cycles. From the figure, we can infer, with the same events, the time required for a Neurotic game companion is longer to feel positive

emotions, but faster for negative emotions than an Extrovert game companion. In addition, the area of positive emotions felt is larger for an Extrovert game companion compares to the Neurotic one. On the other hand, the area of negative emotions felt is larger for Neurotic game companions in comparison to the Extrovert game companions.

Figure 3.6 presents the distribution of social relationship levels of game companion i towards game companion j with the impact of positive emotions on both high Extraversion and high Neuroticism game companions for 1000 cycles. From the figure, we can conclude that, with the same number and quality of interactions, building a social relationship with a Neurotic agent will take extra effort in comparison to building a social relationship with an Extrovert one. The simulation also showed the social relationship value was gradually decreasing when there was no interaction performed between them and was increasing when there is an interaction performed between them. The drop was quite significant as in the simulations, and the interactions were suddenly stopped entirely from having quite positive interactions.

3.5 The ERiSA Framework

In Chapter 2.8.4, we described the required components to build a believable, intelligent VH. We are aiming to replicate human interactive capabilities to perceive information from the external environment, process the information, and react back to the environment. Thus, the main pipeline for the system should include: Sensors to acquire information regarding internal (e.g. the VH's personality and emotions) external (e.g. events and interlocutor's emotions), and indeterminate (e.g. dialogue) environments. The information is then processed based on internal states and rules, resulting in a list of actions proposed. The actions can be verbal (e.g. words), non-verbal (e.g. facial expressions, gestures, prosody, etc.) or a combination of both actions. The actions are then executed and conveyed through the embodiments, be they physical (e.g. robot or audio devices), virtual (e.g. 3D or 2D VHs) or combination of both embodiments. There are some existed integrated systems that are capable of executing the whole pipeline such as (Schroder et al. 2012, Schröder 2010, Hartholt

et al. 2013); hence we will not re-invent the wheel. Instead, we extended the capability of the system and implemented it into a game environment.

3.5.1 The Requirements

As described in the beginning, a framework that capable to perceive information from the external environment, process the information, and react back to the environment is needed to build an emotionally realistic social VH. First, the framework should be able to detect or sense and thus perceive emotions from the user's facial expressions and/or voice prosody. The framework should consist of a component/system that allows the virtual human to understand what the user says. It does not need to understand the entire conversation, as long as the VH be able to create the illusion that they understand what the user says to them. Secondly, the framework should be able to process all the information perceived (user's emotions and utterances) as well as the VH's internal states (the virtual human's emotions, personality and social relationships between the user and the VH). Third, the framework should be able to provide appropriate behaviour feedback and show it through speaker and monitor (the ability to synthesise verbal and non-verbal behaviours).

Broekens et al. (2016) argue that there are six main reasons why emotions models are rare to be implemented in commercial games. Those reasons also provide the requirements of VH implemented in the games (e.g. NPC). Thus, the framework should also be an AI-independent module. Most emotion models are dependent on particular AI. The models and framework thus should be not only dependent on a particular AI module. Chapter 6 describes how the proposed models were implemented to augment existing commercial game characters. Moreover, one of the reasons why emotions models are still not popular in a commercial game is because the players do not demand emotional NPCs Broekens et al. (2016). However, in Chapter 6, we show a preliminary result that the NPCs with the ability to govern emotions increase the players' experiences and immersions to the game, NPC and game story. This could be a good opportunity for commercial game studios to deliver different flavour to enhance the player's experiences and immersions to boost the

game popularity.

3.5.2 The Components

We present a framework for game companions that extends an existing framework, SEMAINE framework (Schröder 2010), by including computational models of *emotions*, *personality* and *social relationships* described in the previous section. We also adopt the framework for a game environment by adding components that manage the gameplay dynamically. The framework consists of modular components including sensing, interpretation, agent, behaviour planning and game components (see Figure 3.7). All the components communicate with each other using XML based messages through ActiveMQ[1], a message-oriented middleware (MOM) as an asynchronous communication system (Schröder 2010). The SEMAINE framework was chosen (despite the Virtual Human Toolkit offers more advantages in some of the components) because at the time the experiments are being conducted, the SEMAINE Framework was the only fully-integrated Virtual Human's Framework API available publicly. Moreover, the SEMAINE Framework provides more advanced non-verbal behaviour recognition components from facial cues and voice prosody features (see Chapter 2.8.4). However, the models and framework proposed in this thesis could technically be implemented to the other SAIBA compliance Virtual Human Framework.

The Sensing Component

The Sensing Component includes modules for Facial Expression Recognition (Action Units using Facial Action Coding System (FACS) and Ekman's six basic emotions (Almaev & Valstar 2013)) from a video, Automatic Speech Recognition and Face Recognition. The Face Recognition module is used to recognise a player's face based on which a game companion can build and interpret a social relationship with a player. The library of known faces is updated dynamically, during interaction between player and game companion. We integrated eMax

[1]http://activemq.apache.org/

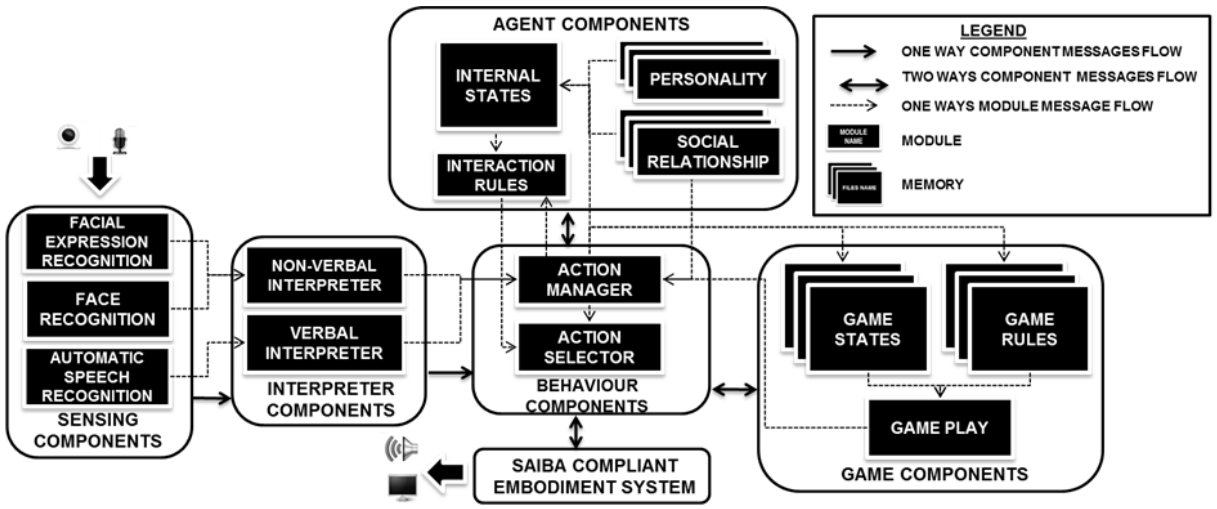


Figure 3.7: Overview of the ERiSA Framework

API (Almaev & Valstar 2013) for expression recognition from facial cues. The Facial Expression Recognition Module recognises six basic emotions (i.e. Happiness, Anger, Sadness, Fear, Disgust and Surprise) from facial cues, in addition to Facial Action Units based on FACS (Ekman 1999*b*). The coding system associates facial expression changes with the atomic facial muscles actions. For example, a smile is associated with the activation of *Zygomaticus major* (AU12) and *Orbicularis oculi, pars orbitalis* (AU6) muscles, along with others facial muscles. The Facial Expression Recognition Module implements a dynamic appearance descriptor with Local Gabor Binary Patterns from Three Orthogonal Planes (LGBP-TOP) proposed by (Almaev et al. 2013). This module returns values of intensity between 0 to 1 for positive emotions (i.e. happiness and surprise) and all Facial Action Units values and -1 to 0 for negative emotions (i.e. anger, sadness, fear, disgust and surprise). All the detected values are stored in the Information States form (see Table 4.7) and sent to the Interpreter Components. The Face Recognition Module implements the Local Binary Patterns Histograms Face Recognizer API from openCV [2]. The module compares the detected face with collected known faces database. If recognised, the module loads all the information about the player and sends all the data to the Interpreter Components along with the confidence value of recognised face. If the face is not recognised, the module will automatically capture several images of the player's frontal face and stored it to the database. Furthermore, an "unrecognised" flag is sent to the Interpreter Components. Finally, we implement the CMU

[2]http://docs.opencv.org/2.4/modules/contrib/doc/facerec/facerec_tutorial.html

Sphinx Speech Recognition (Walker et al. 2004) for Automatic Speech Recognition. All the recognised player utterances are sent to the Interpreter Components to be further processed.

The Interpreter Component

The Interpreter Component receives inputs from The Sensing Components and interprets the received data based on knowledge provided to the system. Player's emotions and action units values from their facial cues are obtained from the Facial Expression Recognition Module. Those values also indirectly update the game companion's internal states in the Agent Components.

Moreover, the components also receive the player's name and all the social relationship information perceived by the game companion (see Figure C.5) from The Face Recognition Module in The Sensing Components. From the Facial Expression Recognition Module, the Non-Verbal Interpreter Module interprets the player's internal states from their facial cues. Currently, the system can interpret the player's emotions from their facial expressions together with the AUs. The components received vectors of numbers indicating the player's emotions and AUs from the Sensing Components. The Interpreter Components then interpret the data received with the rules manually crafted stored in the Information States form. For example: if the data received from the Sensing Components is *$player-states.emotion.happiness* ≥ 0.5, the components interpret the players are currently quite happy or experiencing high positive valence. The Interpreter Components also received all the player's data such as their name, date of birth (if the system has it) and their relationships values. Hence the components will be able to interpret how close the game companion with the player. This information is managed and stored in the Information States form and is also passed to the Agent Components.

Finally, from Automatic Speech Recognition Modules, the Verbal Interpreter Module will spot keywords and patterns. Each keyword or pattern will trigger user intention and update the dialogue state (i.e. topic state) and game state. The Interpreter Components is also responsible for interpreting the turn-taking during the interaction and send the signals to

the Behaviour Components. Listing C.1 exhibits the example of rules for intention and states triggers by patterns or keywords.

The Behaviour Component

The Behaviour Component proposes an agent behaviour based on the information of both verbal and non-verbal interpretation and the Interaction Rules. Table 3.3 shows the example of the interaction rules design for a virtual human in a game scenario. The game companion's actions, in this example, conveyed through their facial expression and utterance is triggered when all the states are matched. In this scenario, the game rules dynamically regulate item price, with the rules based on the player's emotion conveyed through their facial expressions, the game companion's emotion and their relationship level. The price will drop significantly if both player's and game companion's emotions are positive in addition to their positive relationship level and vice versa. The proposed behaviour is sent to a SAIBA compliant embodied system by the Action Selector Module. The Action Manager module then updates all the states and rules including Internal States (e.g. NPC's emotions), Game States (e.g. game level), Interaction Rules, Game Rules, and Social Relationships.

The Agent Components

The Agent Components consist of Memories (see Appendix C) and a Social Relationship module. Memories include the Internal States, Internal Rules and Personality. The Interaction Rules are influenced by the NPC's internal states such as emotions, personality and social relationships. The Social Relationship module manages the relationship status between two entities such as between agent A and player I, player I and player II or agent A and agent B. The Agents Components also consist of dialogue management system and the proposed computational models of emotion, personality and social relationships. The system implements the Information States approach (Traum & Larsson 2003) to dialogue management as the approach proved to be robust and flexible towards a system that has multiple inputs and states (ter Maat & Heylen 2011).

The Game Components

The Game Components consists of a Game Play module and related Memories including Game Rules and the Game States. The Game Play module regulates the gameplay from the Memories. The gameplay changes dynamically based on the NPC's internal states and the player's emotions. The Game Components get inputs from Agent Components, which provide both the player's and game companion's internal states. Hence, it is possible to design game rules based on both the internal and game states. Table 3.3 shows the example of game rules design based on the states. In this example, the item price depends on the game companion's states and the relationship level between the player and the game companion. The same principle can be applied to determine the amount of classified information that game companion will pass to the player based on the states.

3.5.3 Example Game Scenarios Using ERISA

An example in which the social relation is used in a game scenario is as follows: In an RPG game, a player's mission is to convince guard Maximus to let the player pass the gate. The player should do this by building a positive relationship, avoiding conflict. Suppose the player has met Maximus three times before, has a positive relationship with him and Maximus has a high value of extraversion. The player greets the guard with a smile while saying "hello". The player's utterance and facial expression are captured and processed in the Sensing Components and interpreted in the Interpreter Components. The Action Manager updates the values of the social relationship and the guard's internal states. In this case, emotion is set to be more positive, and the value of the social relationship is increased as well. Finally, Maximus reacts based on the information from the Interpreter Components and Interaction Rules. With the framework, the game designer also can build dynamic gameplay that depends on the player's action and the guard's internal states such as emotion, personality and social relationship.

Another example where emotions and social relationships are incorporated in the game

Table 3.3: Game Rules Design Example

Example 1	
Player States	emotion.anger $\leq$ -0.5
Player Intention	ask-wh-question-aboutitemprice
Dialogue States	topic == item
NPC States	emotion.anger $\leq$ -0.5
Relationship Level	affinity $<$ 0.5
Game States	ANY
Game Rules	$item_price = original_item_price * (1 - agentstates.emotion - affinity)$
Action Proposed	faceexp=anger say="I'm not in a good mood, the price is $item_price"
Example 2	
Player States	emotion.anger $\leq$ -0.5
Player Intention	ask-wh-question-aboutitemprice
Dialogue States	topic == item
NPC States	emotion.anger $\leq$ 0.5
Relationship Level	affinity $\geq$ 0.5
Game States	ANY
Game Rules	$item_price = original_item_price * (1 - agentstates.emotion - affinity)$
Action Proposed	faceexp=anger say="What's wrong with you today, $name?, the price is $item_price"
Example 3	
Player States	emotion.happiness $\geq$ 0.5
Player Intention	ask-wh-question-aboutitemprice
Dialogue States	topic == item
NPC States	emotion.happiness $\geq$ 0.5
Relationship Level	affinity $<$ 0.5
Game States	ANY
Game Rules	$item_price = original_item_price * (1 - agentstates.emotion - affinity)$
Action Proposed	faceexp=hapiness say="Hi, I'm in a good mood today, the price is $item_price"
Example 4	
Player States	emotion.happiness $\geq$ 0.5
Player Intention	ask-wh-question-aboutitemprice
Dialogue States	topic == item
NPC States	emotion.happiness $\geq$ 0.5
Relationship Level	affinity $\geq$ 0.5
Game States	ANY
Game Rules	$item_price = original_item_price * (1 - agentstates.emotion - affinity)$
Action Proposed	faceexp=hapiness say="Hey $name, how are you today? The price today is $item_price"

design is when a player is required to retrieve classified information about the when is the next attack will be launch by the enemy from two spies who have a different personality from each other. The player must gain their trust to get information from them. The value and the amount of the information are depending on how close the player with the spies. The player is required to use a different strategy for both of spies, to gain their trust.

3.6 The Evaluation of ERiSA

To explore the potential of the framework and models as well as to collect initial data for interaction rules, we designed two scenarios where the players interacted with the game companions, be they Embodied Conversational Agents (ECAs) or NPCs. The first scenario was a simple social game called *The Smile Game*. The game was inspired by the "Don't Smile" game where two people facing each other and try to make each other laugh by telling jokes or pulling funny facial expressions. To study the interaction rules required when people play this game, we performed an experiment where two people play the game while being video recorded. The video then annotated using the Facial Action Coding System (FACS). *The Smile Game* scenarios are explained over two chapters. Chapter 4 provides the example of the implementation of the proposed framework and computational models in a simple lab setting. While *The Smile Game* scenario in the Chapter 5 intended to explore the interaction between human and game companions in the wild, as well as to evaluate the proposed framework and models.

In the second scenario, we implemented the framework and models for an existing commercial game, Skyrim. In this game, both players and game companions are inhabiting the same world, a virtual world of Skyrim. This scenario aims to explore the player's experience in the game, resulted from the interactions with the game companions, who were imbued with social and emotional skill by the ERiSA Framework together with the computational models. The Skyrim Game scenario is discussed comprehensively in the Chapter 6. Table 3.4 illustrates the overview of the framework, computational models and the game design

implemented in each study. To continuously improve the quality of the interactions between players and the game companions, the system's design (framework, computational models and game design) has evolved throughout the series of the experiments. There are two major types of frameworks used in the experiments: Full and Simplified. Full means in the experiment, all the components in the framework are fully functioning, though, in some experiments, some modules were replaced by other modules (e.g. ASR module in the Sensing Components replaced by another input module, The Text module from the keyboard). Simplified framework only implements some of the Components from the proposed framework, as the other Components are not necessarily required in the experiment (see Chapter 6). Moreover, they are also two versions of Computational Models of Emotions, Personality, and Social Relationships. The first one was described in Section 3.2, while the second one was comprehensively illustrated in Section 3.3. Finally, there are two version of the game design in the Smile Game (in contrary, only one version for the Skyrim Game), the first one (SMLGD V.1) is described in this chapter, while the enhanced version (SMLGD V.2) is illustrated in the Chapter 5.

Table 3.4: Versioning Overview

Version	**Framework**	**Models Ver.**	**Game Design Ver.**
Smile Game V.1 (SML V.1)	Full	1	SMLGD V.1
Smile Game V.2 (SML V.2)	Full	2	SMLGD V.1
Smile Game V.3 (SML V.3)	Full	2	SMLGD V.2
Skyrim Game V.1 (SKR V.1)	Simplified	2	SKRGD V.1
Workshop V.1 (WRK V.1)	Full	2	#NA

Table 3.5 describes the overview of the user study design of all studies conducted over three years with a total of 117 participants in 217 interactions with the system. The first study was intended to collect behaviour data from human-human interaction when playing a simple social game called *The Smile Game*. Ten adult participants played the game in pairs, resulting in eighteen sessions of interaction. The second study was a pilot study of *The Smile Game*, involving 16 adults interacted with the system in a lab setting in a total of 64 interactions. The pilot study aimed to evaluate the early version of computational models and the integrated system functionality as well as to collect additional behaviour data for conversation and game interaction. Figure 3.8 illustrates the setting of the studies. The

third and fourth studies involved a total of twenty adult participants in a total of twenty-one interactions in a *semi-public* space setting. The first to third settings were involved facial expressions and speech as the inputs in the interactions, while in the fourth set, a joystick was introduced as an additional input besides the player's facial expressions during the interaction in the game.



Figure 3.8: Settings of The Smile Game: Pilot (a)WoZ & Auto(b) Mayfest(c) & Skyrim(d)

The purpose of the studies was to evaluate the second version of computational models and the integrated system as well as to collect more behaviour data for the game companions. The fifth and sixth studies were designed for public interaction. A total of 54 participants in 58 interactions were playing with the system. The seventh study was designed to evaluate the computational models and some parts of the integrated system in an existing commercial game. Finally, a character design workshop was conducted to explore the potential of the computational models and the integrated system. In this study, we were aiming to investigate how professional game designers would implement our models and system in their game design. The Smile Game evaluations are thoroughly described in this chapter and Chapter 5. The first and second studies are presented in Section 4.5, while third and fourth are explained in Section 5.2. Moreover, the fifth and sixth studies are described in Section 5.5 and The Skyrim Game study is illustrated in Section 6. Finally, the design workshop is presented in the Section 7.

Table 3.5: User Studies Overview

Study	Version	Participants	# Int.	Setting	Aim	Chapter
H - H	#NA	10 (A)	18	Lab	Data Collection	4
Pilot	SML V.1	16 (A)	64	Lab	System + Models Evaluation Data Collection	4
WoZ	SML V.2	8 (A)	9	Semi-Public	System + Models Evaluation Data Collection + Interaction in the Wild	5
Auto	SML V.2	12 (A)	12	Semi-Public	System + Models Evaluation Data Collection + Interaction in the Wild	5
Mayfest	SML V.3	46 (E)	50	Public	System + Models Evaluation + Interaction in the Wild	5
OpenDays	SML V.3	8 (E)	8	Public	System + Models Evaluation + Interaction in the Wild	5
Skyrim	SKR V.1	15 (A)	60	Lab	Models Evaluation	6
Workshop	WRK V.1	2 (A)	2	Lab	System + Models Exploration	7

[This page intentionally left blank]

Chapter 4

Building Game Companions with ERiSA

4.1 Introduction

This chapter presents the effort to build game companions using the proposed framework and models. The aims are to explore further the potential of the framework and models deployed in a simple interaction system, and to collect the interaction rules when people were interacting with each other given a particular scenario. A simple social game called *The Smile Game* is designed to be the testbed for the proposed framework and models evaluation. The game was chosen because the game allows the players who are playing it eliciting rich expressions and interaction. This chapter also describes the initial design of game companions using The Smile Game V.1 (see Table 3.4). Section 4.2 describes the data collection to craft the interaction rules based on the behaviours that people would exhibit in the game. Section 4.5 shows the behaviour rules crafted from the data collection, annotation, and analysis. Moreover, Section 4.3 illustrates the initial game design for *The Smile Game*. The game then evaluated as a pilot study to serve as a foundation of the other experiments in the affective loop experiences system. The experimental settings are described in section 4.5, and the results are shown in section 4.6 and concluded in section 4.7



Figure 4.1: The Recording Settings

4.2 Data Collection

There are two categories of behaviours that the game companions would exhibit in this type of scenario: the behaviour when playing the game and the behaviour when the player and the game companions having a conversation. To study what type of behaviour people playing the Smile Game would exhibit, we performed an experiment in which two people at a time play the game while being video recorded. This will allow us to model typical behaviour and replicate it in the game companions. Moreover, to learn the dialogue rules in the conversation, we annotated the transcripts of the conversations between human and agent (controlled by an operator) interaction from The SEMAINE Database Corpus (McKeown et al. 2012), with Dialogue Acts annotation (Austin 1975).

4.2.1 Human-Human Interaction Data Collection

We recruited 10 participants, all of whom are students of The University of Nottingham (5 male; 6 Asian, 4 Caucasian). We conducted five recording sessions, with a different number of games played per session resulting in a total of 18 interactions recorded.

During the game, three cameras were recording both visual and audio information. The first and second camera which each records a player from the front while the third camera records a profile perspective of both players (see Figure 4.2 and Figure 4.1). The images are used for facial expressions analysis and the audio recording for discourse analysis (i.e. presence of conversation, and who spoke first). Detail of the information of the recording is shown in Table 4.1. Column S represents the recording session, for example, 1 indicates Session no. 1. Column VL(s) indicates the length of the video clip in second. Column #GAME demonstrates the number of game played each session, for example, the value 2 in the first row means there are two games in the Session no.1. RELATIONSHIP column shows the relationship between the player. The relationship was classified into two categories: close and very close relationship, based on how long they have known each other and their social attitude to each other. The column GAME indicates the game session number, for example, the value G1 in the first row indicates Game 1 in the Session 1. Column GL(s) shows the length of the game in second. Finally, the last column indicates whether any verbal behaviours were performed during the game.

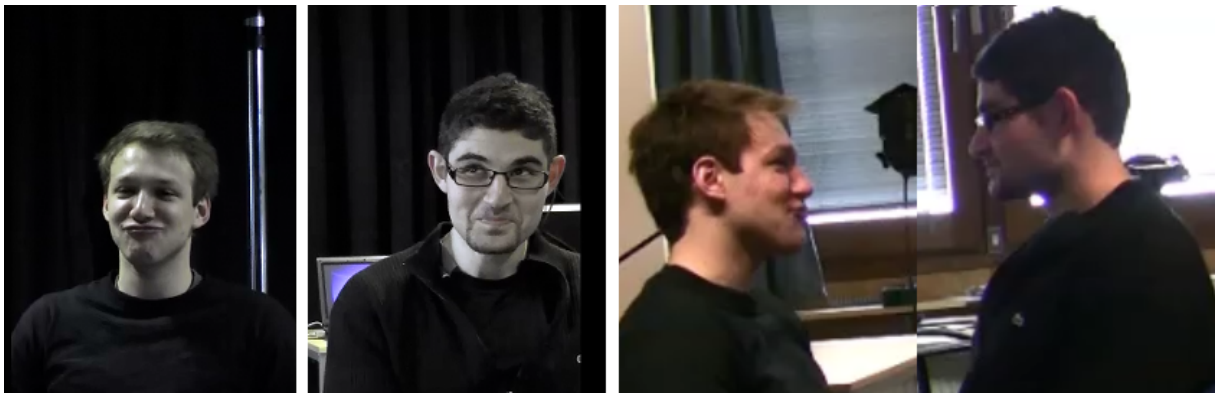


Figure 4.2: The Smile Game Recording

We annotated the recordings using the Facial Action Coding System (FACS) (Ekman et al. 2002*a*) for the players' facial expressions (see Figure 4.3) and created a transcript of the players' utterances[1]. Furthermore, we classified them into two categories: Attack and Defend. Attack is when the player pulled their funny faces or telling jokes to make the other player laugh. Figure 4.3 (c) shows a player's facial expressions when he was attacking his

[1]The utterances in session 4 were in Indonesian; hence, we did a translation to English

Table 4.1: Smile Game Recording Summary

S	VL(s)	#GAME	RELATIONSHIP	GAME	GL(s)	CHAT
1	34	2	Workmate(close)	G1	13	no
				G2	20	no
2	65	3	Classmate(close)	G1	8	no
				G2	18	no
				G2	13	no
3	127	9	Classmate(Close)	G1	127	no
4	1221	9	Friend(very close)	G1	2	no
				G2	10	no
				G3	19	no
				G4	26	yes
				G5	37	yes
				G6	13	yes
				G7	143	yes
				G8	23	yes
				G9	361	yes
5	105	3	Friend(very close)	G1	15	yes
				G2	19	yes
				G3	107	yes

opponent by pouting his face. Meanwhile, Defend is when the player is trying to mask their smile with other expressions. When in Defend position, the player can either stay in that position or counterattack to the other player. The distinct difference between Attack and Defend is that the trace of a smile is clearly visible with medium to high intensity when defending. Figure 4.3 (a) illustrates a Defend position performed by a participant. The participant tried to hold his smile by slightly pressing his lip (AU 24B) and looking to the other direction to avoid eye contact with his opponent. Finally, we selected utterances and facial expressions that were used for attacks to make the opponent laugh (see Table 4.2) as well as for defend to mask the player's smile with other expressions. These attack and defend behaviours were then implemented in the game companions. Table 4.3 lists all observed attack and utterances.

Attack Behaviours

There was a total of 76 sequences of attack where most of them (54%) were performed through facial expressions, head movements, and their combination. Almost two third of

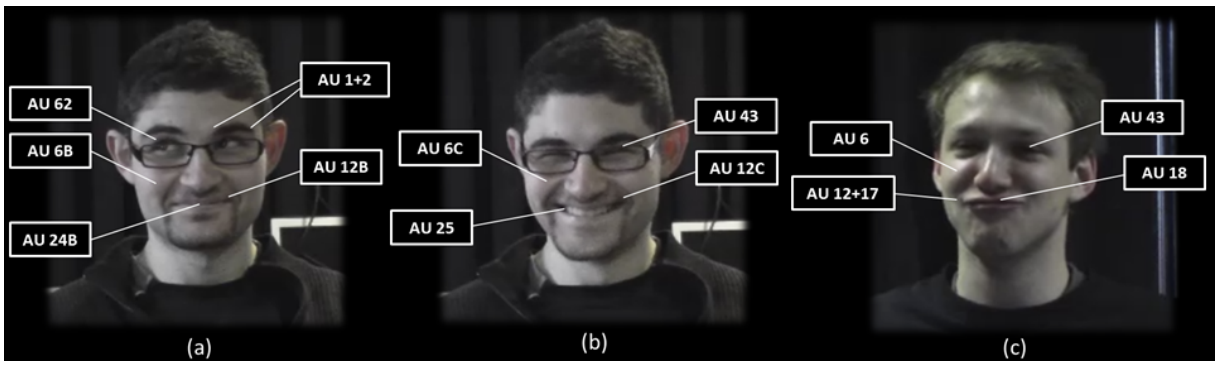


Figure 4.3: Video Annotation using FACS. (a) Defend Behaviour, (b) Smile Behaviour, (c) Attack Behaviour

the participants were only attacking without jokes. In contrast, two participants were using jokes/utterances (76%) for their main attack and defend strategy. Table 4.2 presents all the facial expressions, head movements, and their combination attacks performed in the games.

From the recording, we categorised the players' jokes/utterances into random (where the players are just talking nonsense), commenting on their opponent appearance (either praising or insulting), and offence category (see Table 4.3). Most of the jokes performed in the recording can be only implemented if the players are known each other. Some jokes might be considered rude if the players are new to each other. Such as insulting the opponent's appearance. However, it would be a good attack strategy if the players are relatively close to each other. Some of the jokes performed were sounded meaningless (such as Eiffel tower, and I am sleepy now). However, it could be a good strategy when they also perform random jokes. Such as: when a player said "Once upon a time" and the opponent paused for a couple of seconds and replied "Eiffel tower".

There was a total of 19 distinct FACS combinations as well as 14 utterances performed during the games. We push all these facial expressions, head movement attacks and utterances to the game companions' attack vocabulary. In addition, we added twenty additional funny sentences from various sources (See Appendix 1).

Table 4.2: Attack Behaviour (Non-Verbal)

FACS	MODALITY
AU15	face
AU24	face
AU27	face
AU23 + AU26	face
AU4 + AU17	face
AU18 + AU26	face
AU6 + AU18 + AU62	face
AU1 + AU2	face
AU1 + AU2 + AU17	face
AU1 + AU2 + AU27	face
(AU62 + AU63) ->AU63 ->(AU61 + AU63) ->AU63	face
AU51 ->AU52 ->AU51 ->AU52	head
AU56 + AU61 + AU63	head
AU54 + AU61 + AU63	head
AU54 ->AU53 ->AU54 ->AU53 ->AU54 ->AU53	head
AU1 + AU2 + + AU27 + AU55 ->AU56	face + head
AU1 + AU2 + + AU27 + AU51 ->AU52 ->AU51 ->AU52	face + head
AU24 + AU43 + AU55	face + head
AU1 + AU2 + AU17 + AU55 ->AU56	face + head

Defend Behaviours

Defend behaviours are classified into two categories: passive and active. Defend-passive is a behaviour shown when the player try to hold their laugh (e.g. AU6 + AU12 + AU24 or AU12 + AU6 + AU61) (see Figure 4.3 (a)) while Defend-active is a behaviour shown when the player try to mask their smile with attacking behaviour (e.g. AU6 + AU12 + AU24 + AU1 + AU2 or AU6 + AU12 + AU18 + AU24). In the recording, most of the players consistently displayed a Defend-passive behaviour before they burst into a peal of laughter. While some players exhibited Defend-active behaviour to counter-attack their opponent. There was a total of sixty sequences of defend behaviours displayed in the recording, where most of them (57%) were Defend-passive behaviour. There were also some utterances that can be applied for defend strategy (see Table 4.3). All distinct FACS combination in Table 4.4, as well as seven distinct utterances performed for defend strategy in the recording, were pushed to the game companions' defend vocabulary.

Table 4.3: Attack and Defend Behaviour (Utterances)

UTTERANCES	UTTERANCES
Once upon a time	random
Eiffel tower	random
I am sleepy now	random
Hi mate	random
You've got pretty eyes	appearance (positive)
You curl your hair!	appearance (positive)
You need to wash your clothes!	appearance (negative)
You need to wash your hair	appearance (negative)
Did you brush your teeth today?	appearance (negative)
you need to do eyebrows	appearance (negative)
Running out things to say	offence
You are not as funny as you think	offence
Dying to laugh, aren't you?	offence
Come on laugh!	offence
not a bit	defend
yes or yes I do(did)	defend
maybe	defend
cheers mate	defend
so do you	defend
why aren't you laughing	defend
are you trying to hurt me now?	defend

Smile Behaviours

Most of the players (60%) were laughing as soon as the game started. Some people were laughing immediately after pulling their funny faces, as they thought about what they were doing was ridiculous. There were 86 sequences of Smile behaviour shown in the recording. However, to simplify the game companion's smile behaviour, only the combination of AU6, AU12 and their intensity were implemented into the game companion's smile behaviour vocabulary. Figure 4.3 (b) illustrates the player's facial expression when the player was laughing after the opponent performed an attack behaviour to him.

The Game Companion's Strategy

The annotation from the interaction video provides four behaviours to perform when playing this type of game. The interaction video also contributes to a pattern of behaviour shown in the game. This pattern was then implemented into the game companion's strategy to win the

Table 4.4: Defend Behaviour (Non-Verbal)

FACS	MODALITY	CATEGORY
AU1 + AU2 + AU6 + AU12	face	Defend-active
AU6 + AU12	face	Defend-passive
AU6 + AU12 + AU18	face	Defend-passive
AU6 + AU12 + AU18 + AU24	face	Defend-active
AU6 + AU12 + AU18 + AU43	face	Defend-passive
AU6 + AU12 + AU23	face	Defend-passive
AU6 + AU12 + AU24	face	Defend-passive
AU6 + AU12 + AU24 + AU43	face	Defend-passive
AU6 + AU12 + AU24 + AU43	face	Defend-passive
AU6 + AU12 + AU24 + AU53 ->AU54	face + head	Defend-actice
AU6 + AU12 + AU26 + AU52	face + head	Defend-active
AU6 + AU12 + AU27	face	Defend-active
AU6 + AU12 + AU27 + AU43	face	Defend-active
AU6 + AU12 + AU43	face	Defend-passive
AU6 + AU12 + AU43 + AU52	face + head	Defend-passive
AU6 + AU12 + AU43 + AU52 + AU62	face + head	Defend-passive
AU6 + AU12 + AU45	face	Defend-passive
AU6 + AU12 + AU51	face + head	Defend-passive
AU6 + AU12 + AU56	face + head	Defend-passive
AU6 + AU12 + AU56 + AU58	face + head	Defend-passive
AU6 + AU12 + AU62 + AU63	face	Defend-passive
AU6 + AU12 + AU43 + AU51 ->AU52	face + head	Defend-active

game in *The Smile Game*. We adopt a finite state's approach to the game companion's action were the four behaviours will be the states. The game companion's smile threshold calculated from the AUs perceived from the player's facial expressions and the player's internal states will be the triggering events for a transition to the other states. Figure 4.4 illustrates the finite states of the game companion's behaviours in the game. When the game starts, the game companion will always start the attack. If the value of the game companion's urge to smile at time t (S_t, see 4.3.1) hits the first game companion's threshold for smiling (θ_{S1}, see 4.3.1), the game companion will display modest cues for smiling and enter Defend behaviour mode. A simple random number (Rnd_t) will decide whether the game companion will stay on the Defend-passive mode or mask their smile with an attack in the Defend-active mode at time t. To make it more apparent, the game companion informs the player that the player is almost making them laugh in addition to the small smile cues conveyed through the NPC's facial expressions. When the value of the game companion's urge to smile hits the second game companion's threshold for smiling (θ_{S2}, see 4.3.1), the game companion will break into

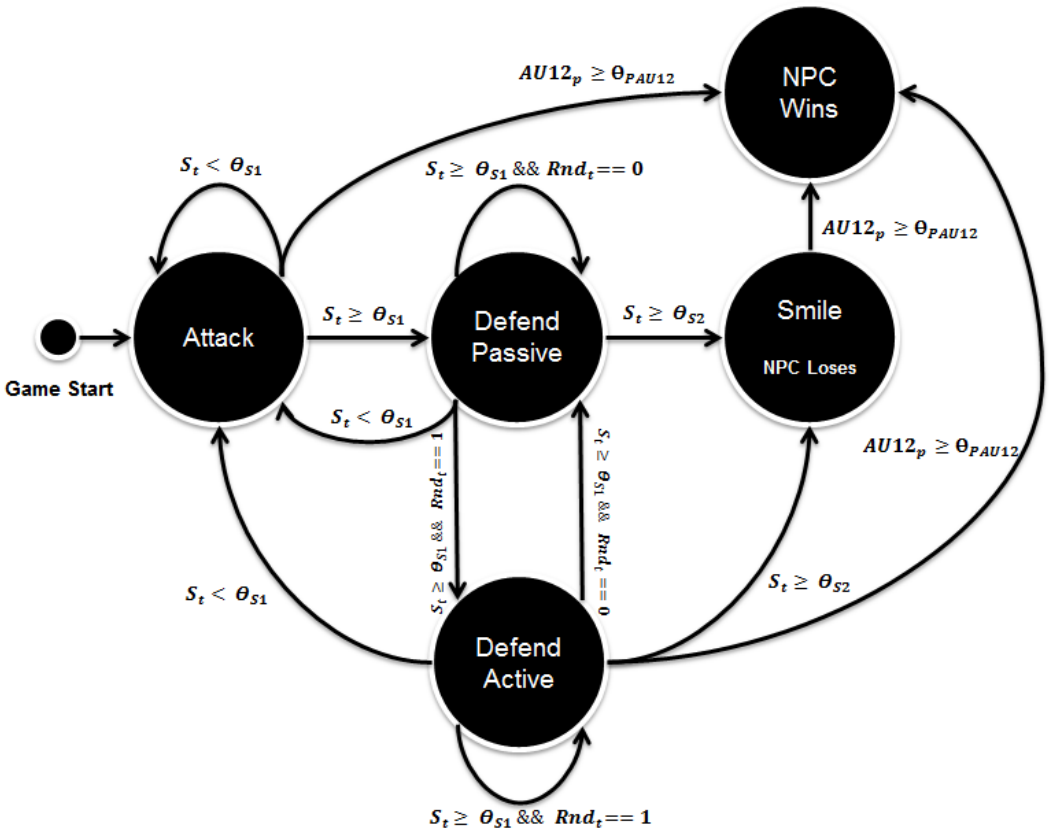


Figure 4.4: Game Companion's Strategy Finite State

a smile, and the game ends. Equally, if the player's AU12 value perceived from their facial cues ($AU12_p$) reached a specific value of threshold (θ_{PAU12}), the player loses, and the game ends. The threshold θ_{PAU12} can be adjusted dynamically upon the game level.

4.2.2 The SEMAINE Database Corpus

To learn the dialogue rules in the conversation, we annotated The SEMAINE Database Corpus (McKeown et al. 2012) using Dialogue Act annotation scheme (Austin 1975). Since we only focus on two out of five traits (i.e. extraversion and neuroticism) of personality proposed in the models, we only annotated the interaction with two SEMAINE characters (i.e. Poppy and Spike) those represent the two of personality traits. Forty transcripts of the conversations were annotated resulting 2100 annotations, composed in 54 dialogue acts. Table 4.5 shows the overview of the Dialogue Acts classification from the SEMAINE Database Corpus.

Table 4.5: Dialogue Acts Classification Overview

Dialogue Act	#	**Dialogue Act**	#
acknowledgement	89	statement-nonopinion-aboutplace	14
agreement	32	statement-nonopinion-aboutplace-neg	8
answer-wh-question-aboutagent	49	statement-nonopinion-aboutweather-neg	16
answer-wh-question-aboutfamily	26	statement-nonopinion-aboutweather-pos	38
answer-wh-question-aboutfriend	32	statement-opinion-aboutagent	144
answer-wh-question-aboutholiday	22	statement-opinion-aboutagent-neg	140
answer-wh-question-aboutotheragent	16	statement-opinion-aboutbook	18
answer-wh-question-aboutuser	94	statement-opinion-aboutfamily	8
answer-wh-question-aboutwork	16	statement-opinion-aboutfamily-neg	1
answer-wh-question-name	12	statement-opinion-aboutfriend	76
answer-yes-no-question	26	statement-opinion-aboutfriend-neg	3
appreciation	4	statement-opinion-aboutfriend-pos	2
ask-wh-question-aboutagent	49	statement-opinion-abouthobby	18
ask-wh-question-aboutfamily	26	statement-opinion-aboutholiday	122
ask-wh-question-aboutfriend	32	statement-opinion-aboutholiday-neg	26
ask-wh-question-aboutholiday	22	statement-opinion-aboutholiday-pos	40
ask-wh-question-aboutotheragent	16	statement-opinion-aboutmovie	8
ask-wh-question-aboutuser	94	statement-opinion-aboutotheragent	8
ask-wh-question-aboutwork	16	statement-opinion-aboutplace	120
ask-wh-question-name	12	statement-opinion-aboutplace-neg	6
ask-yes-no-question	26	statement-opinion-aboutplace-pos	2
disagreement	123	statement-opinion-aboutpolitics	25
goodbye	70	statement-opinion-aboutshopping	62
greeting	142	statement-opinion-aboutuser	3
no answers	1	statement-opinion-aboutuser-neg	8
statement-nonopinion-aboutagent	28	statement-opinion-aboutwork	39

Table 4.5: Dialogue Acts Classification Overview

Dialogue Act	#	**Dialogue Act**	#
statement-nonopinion-aboutholiday	68	statement-opinion-aboutwork-neg	2

The annotation format we used in this interaction video is: ***[dialogue acts]-[topic]-[valence]***[2]. For example: the annotation of statement-nonopinion-aboutweather-neg indicates that the interlocutor was expressing their non-opinion statement (the dialogue acts) about the weather (the topic) and conveying a negative emotions (the valence) verbally or non-verbally (e.g. facial expressions or audio pitch) or the annotation of statement-nonopinion-aboutweather shows that the interlocutor was expressing the statement casually with neutral emotions. Table 4.6 shows the pairs of the utterance in the SEMAINE Database Corpus. We adopted adjacency pairs to determine pairs of the utterances. With the pairs, we can model the answer utterance given a particular utterance from the player. There are sixty-one pairs in the annotation with the most common pair appeared in the interaction is statement-opinion about a particular topic. Some Dialogue Acts are paired with more than one other Dialogue Acts. For example, in the database, the Dialogue Acts *statement-nonopinion-aboutagent* can be replied with the utterances that belong to the Dialogue Acts *disagreement*, *acknowledgement* or *statement-nonopinion-aboutagent*.

Table 4.6: Adjacency Pairs

Interlocutor 1	**Interlocutor 2**	**Count**
ask-wh-question-aboutagent	answer-wh-question-aboutagent	49
ask-wh-question-aboutfamily	answer-wh-question-aboutfamily	26
ask-wh-question-aboutfriend	answer-wh-question-aboutfriend	32
ask-wh-question-aboutholiday	answer-wh-question-aboutholiday	22
ask-wh-question-aboutotheragent	answer-wh-question-aboutotheragent	16

[2][optional]

Table 4.6: Adjacency Pairs

Interlocutor 1	**Interlocutor 2**	**Count**
ask-wh-question-aboutuser	answer-wh-question-aboutuser	94
ask-wh-question-aboutwork	answer-wh-question-aboutwork	16
ask-wh-question-name	answer-wh-question-name	12
ask-yes-no-question	answer-yes-no-question	26
disagreement	disagreement	6
goodbye	goodbye	35
greeting	greeting	71
statement-nonopinion-aboutagent	disagreement	8
statement-nonopinion-aboutagent	acknowledgement	4
statement-nonopinion-aboutagent	statement-nonopinion-aboutagent	8
statement-nonopinion-aboutholiday	statement-nonopinion-aboutholiday	21
statement-nonopinion-aboutholiday	acknowledgement	26
statement-nonopinion-aboutplace	acknowledgement	14
statement-nonopinion-aboutplace-neg	acknowledgement	4
statement-nonopinion-aboutplace-neg	agreement	4
statement-nonopinion-aboutweather-neg	acknowledgement	6
statement-nonopinion-aboutweather-neg	statement-nonopinion-aboutweather-neg	5
statement-nonopinion-aboutweather-pos	acknowledgement	14
statement-nonopinion-aboutweather-pos	statement-nonopinion-aboutweather-pos	12
statement-opinion-aboutagent	statement-opinion-aboutagent	44
statement-opinion-aboutagent-neg	disagreement	84
statement-opinion-aboutagent-neg	statement-opinion-aboutagent	56
statement-opinion-aboutbook	agreement	8
statement-opinion-aboutbook	statement-opinion-aboutbook	5
statement-opinion-aboutfamily	acknowledgement	8
statement-opinion-aboutfamily-neg	no answers	1

Table 4.6: Adjacency Pairs

Interlocutor 1	**Interlocutor 2**	**Count**
statement-opinion-aboutfriend	statement-opinion-aboutfriend	38
statement-opinion-aboutfriend-neg	disagreement	3
statement-opinion-aboutfriend-pos	acknowledgement	2
statement-opinion-abouthobby	statement-opinion-abouthobby	9
statement-opinion-aboutholiday	statement-opinion-aboutholiday	61
statement-opinion-aboutholiday-neg	statement-opinion-aboutholiday-neg	11
statement-opinion-aboutholiday-neg	disagreement	4
statement-opinion-aboutholiday-pos	statement-opinion-aboutholiday-pos	20
statement-opinion-aboutmovie	statement-opinion-aboutmovie	3
statement-opinion-aboutmovie	agreement	2
statement-opinion-aboutotheragent	appreciation	2
statement-opinion-aboutotheragent	statement-opinion-aboutotheragent	3
statement-opinion-aboutplace	statement-opinion-aboutplace	58
statement-opinion-aboutplace	acknowledgement	4
statement-opinion-aboutplace-neg	acknowledgement	4
statement-opinion-aboutplace-neg	agreement	2
statement-opinion-aboutplace-pos	appreciation	2
statement-opinion-aboutpolitics	statement-opinion-aboutpolitics	10
statement-opinion-aboutpolitics	disagreement	2
statement-opinion-aboutpolitics	agreement	3
statement-opinion-aboutshopping	statement-opinion-aboutshopping	28
statement-opinion-aboutshopping	agreement	6
statement-opinion-aboutuser	acknowledgement	3
statement-opinion-aboutuser-neg	disagreement	8
statement-opinion-aboutwork	agreement	5
statement-opinion-aboutwork	statement-opinion-aboutwork	16

Table 4.6: Adjacency Pairs

Interlocutor 1	**Interlocutor 2**	**Count**
statement-opinion-aboutwork	disagreement	2
statement-opinion-aboutwork-neg	agreement	2

4.3 Game Design

We implemented the proposed framework and the first version of the computational models (see 3.2) to create a simple game called *The Smile Game*. The main purpose of this project is neither to detect nor to synthesise laughter but to implement and explore the potential of the proposed framework and models that consider the relationship between emotions, personality and social relationships to enhance the game companion's believability and quality of player interaction in the game. Recent work in detecting and synthesising laughter can be found in the Ilhaire project [3] (Urbain et al. 2013). In *The Smile Game* scenario, we replace one of the players with an ECA. We use two SEMAINE characters, Poppy and Spike (McKeown et al. 2012) as the game companions. Poppy has a high extraversion and low neuroticism personality ($\mathcal{P}_E = 0.5, \mathcal{P}_N = -0.1$), while Spike has a high neuroticism and low extraversion personality ($\mathcal{P}_E = 0.2, \mathcal{P}_N = 0.5$). This game is considered to be a good case study since the gameplay is simple yet elicits rich non-verbal interactions between the game companions and the player.

4.3.1 The Smile Game Flow

An overview of the game flow is shown in Figure 4.5. When the system starts, it searches for a face. When a face is detected, the game companion tries to recognise the detected face by comparing it with the database of faces. Based on this information, the game companion

[3] http://www.ilhaire.eu/

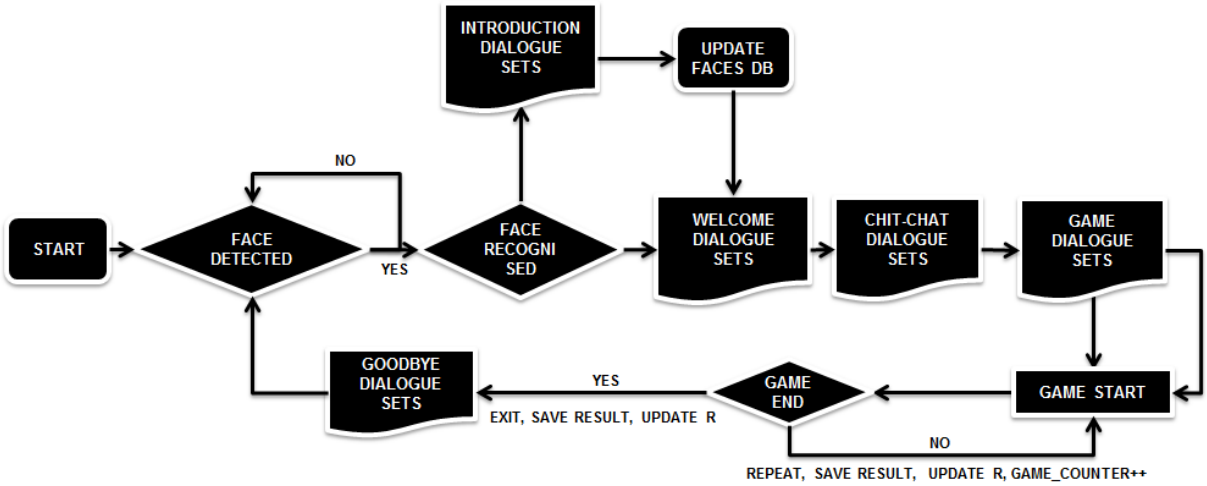


Figure 4.5: The Smile Game flow

greets the player using the Welcome Dialogue Sets (i.e. dialogues with dialogue acts tag *greeting*). If the face is not recognised, the game companion will ask for the player's name and will initialise a new entry in the face library as well as a relationship value $\mathcal{R}_t$. If the system recognises the face, it searches for the corresponding relationship value $\mathcal{R}_t$ and updates it. Afterwards, the game companion starts a conversation using common topics such as the weather or news using the Chit-Chat Dialogue Sets (e.g. dialogues with dialogue acts tag *statement-opinion-aboutweather*, *statement-opinion-aboutpolitics* or *statement-opinion-aboutholiday*). After a certain period of time, the game companion will ask whether the player would like to play a game with the game companion. If so, the game will commence and continues until someone smiles. The player can also ask the game companion to play a game with them. When the game ends, the result of the game is saved, in addition to updating the value of $\mathcal{R}$. Both game companion and player can ask to repeat the game if the player does not want to repeat the game, the game companion closes the interaction using its Goodbye Dialogue Sets (i.e. dialogues with dialogue acts tag *goodbye*).

In this game, we categorise relationship between the player and the game companion into two groups of pairs in addition to the "New" category. Along with the game companion's emotions and personality, the social relationships level modifies the answers given by the game companions based on the models proposed in Section 3. The pairs are: "friend" vs "rival" and "ally" vs "enemy". For example: When the default utterance to reply a greeting

(category = *greeting*) from the player is *"Hello!"*, the answer to reply a greeting (category = *greeting-posLowRelationship*), when their social relationship level is a friend, is *"Hello, $name!, shall we play a round of game? I think I can beat you this time!"* and a reply such as *" It is you again? What do you want from me?"* is provided when the relationship level they have is a rival (category = *greeting-negLowRelationship*). Moreover, the models of emotions, personality and social relationships also influence the game companion's urge to smile.

The dialogue states in the system are categorised into three categories: Greeting, Chit-Chat and Goodbye, in which each state has its own dialogue sets derived from the data collection described in Chapter 3 with additional behaviour data from the studies presented in this chapter. Additionally, we extended the dialogue sets with dialogue related to the game such as utterances to explain about the game (Dialogue Acts category: statement-nonopinion-aboutthegame), utterances to explain the game rules (Dialogue Acts category: *statement-nonopinion-about*) or utterances to ask the player if they are ready to play (Dialogue Acts category: *ask-yes-no-question-readytoplay*). Furthermore, the game companion also has a dialogue sets to portray their background stories related to the game which is described in the Appendix C.

In addition to the emotions, personality and social relationships from subsection 3.2, we define a model to determine when the game companion will smile. The game companion's urge to smile at time t S_t is derived from the model of emotion E_t from section 3.2. The difference is that while the model of emotion E_t considers all the possible player's verbal and non-verbal actions, the model of smile S_t only considers the possible player's verbal and non-verbal actions that could trigger the game companion to smile. Currently, the same rules used by the game companion to attack an opponent will incur an increase in S_t when observed by the game companion.

To model the dynamic properties of the game companion's smile in a simple yet convincing way, we again included a decay function to our model. This way the game companion's urge to smile will diminish over time unless reactivated. Similar to the emotion model, the value

of the decay rate r on this model is also set based on the game companion's personality. The game companion has two thresholds of a smile. The first threshold θ_{S1} triggers small cues for a smile in its facial expression (e.g. AU12B), where the second threshold θ_{S1} activates a clear sign of a smile on its facial expressions (e.g. AU12D + AU6C). Both smile thresholds are regulated by the game companion personality and the game level. The game companion's urge to smile S_t denotes in the Equation 4.1, where I_t represents the game companions emotion intensity described in Section 3.

$$S_t = S_{t-1} \times e^{-rt} + I_t \tag{4.1}$$

4.3.2 The Game Storyboard

Case Study 1: First Meeting with Poppy

1. **John sits in front of the system.**

   The system detects a face.

2. **The system introduces itself and explains what it can do.**

   When a face is detected, the system starts to introduce itself and explain to the user what the system can do.

3. **The system asks John: "Who would you like to speak with?"**

   After introducing and explaining itself, the system asks the user who is the game companion that user want to speak with. The system introduces the game companions: Poppy and Spike. In addition, it describes Poppy and Spike's personality as well.

4. **John asks to speak with Poppy**

   The system registered Poppy as the game companion, and then the screen changes into Poppy's face.

5. **John meets with Poppy for the first time.**

   The system once again tries to detect a face. When the system detects a face, it starts

facial recognition modules and tries to recognise John. If John is not recognised, the system then asks John's name.

6. **Poppy asks John's name by typing it.**

   The system captures John's face ten times and stores the images as face database. John then types his name using a keyboard. Keyboard system is used to avoid any mistakes in the user's name spelling in speech recognition modules.

7. **Poppy welcomes John and starts the conversation.**

   Poppy welcomes John to the Smile Game and asks John about the newest news or weather. Poppy has high extraversion and low neuroticism personality; Hence she tends to express and perceive positive emotions such as happiness. On the other hand, Spike has high neuroticism and low extraversion personality, so he tends to express and perceive negative emotions such as anger.

Case Study 2: Speak with Poppy

1. **John sits in front of the system.**

   The system detects a face.

2. **The system introduces itself and explains what it can do.**

   When a face is detected, the system starts to introduce itself and explain to the user what the system can do.

3. **The system asks John: "Who would you like to speak with?"**

   After introducing and explaining itself, the system asks the user who is the game companion that user want to speak with. The system introduces the game companions: Poppy and Spike. In addition, it describes Poppy and Spike's personality as well.

4. **John asks to speak with Poppy**

   The system registered Poppy as the game companion, and then the screen changes into Poppy's face.

5. **Poppy says hello and asks John about the newest news**

   The system will randomise the topics. There are two topics that currently available, news and weather. The system then selects an utterance randomly in the chit-chat dialogue sets based on the topic selected.

6. **John says he has not heard any newest news and ask Poppy to tell him the details about the news.**

   When the user says something, the system starts speech recognition modules to recognise the user's utterance. The utterance then analysed, and the system selects the best action.

7. **Poppy tells John about the newest news, and asks John's opinion regarding the news.**

   The system picks an utterance in the chit-chat dialogues sets based on news topic randomly. Currently, the news utterance is stored statically or pre-defined.

8. **John says it is good news to Poppy with a smile in John's face**

   The system starts speech recognition modules to recognise the user's utterance as well as the user's facial expression.

9. **Poppy says "Indeed it is good news. I am glad you happy with the news"**

   Poppy understands John's utterance and reads John's smile. Poppy tends to express and perceive positive emotions such as happiness; hence she is happy when John happy.

10. **Poppy asks John to play a game.**

    The system moves to the next state, the game state where John can play a smile game with Poppy.

Case Study 3 : Speak with Spike

1. **John sits in front of the system.**

   The system detects a face.

2. **The system introduces itself and explains what it can do.**

   When a face is detected, the system starts to introduce itself and explain to the user what the system can do.

3. **The system asks John: "Who would you like to speak with?"**

   After introducing and explaining itself, the system asks the user who is the game companion that user want to speak with. The system introduces the game companions: Poppy and Spike. In addition, it describes Poppy and Spike's personality as well.

4. **John asks to speak with Spike**

   The system registered Spike as the game companion, and then the screen changes into Spike's face. Let's says John has met Spike before; hence Spike recognises John.

5. **Spike says hello and asks John about the newest news**

   The system will randomise the topics. There are two topics that currently available, news and weather. The system then selects an utterance randomly in the chit-chat dialogue sets based on the topic selected.

6. **John says he has not heard any newest news and ask Spike to tell him the details about the news.**

   When the user says something, the system starts speech recognition modules to recognise the user's utterance. The utterance then analysed, and the system selects the best action.

7. **Spike tells John about the newest news, and asks John's opinion regarding the news.**

   The system picks an utterance in the chit-chat dialogues sets based on news topic randomly. Currently, the news utterance is stored statically or pre-defined.

8. **John says it is good news to Spike with a smile in John's face**

   The system starts speech recognition modules to recognise the user's utterance as well as the user's facial expression.

9. **Spike says "Yeah, life sucks."**

   Spike understand John's utterance and reads Johns smile. Spike tends to express and perceive negative emotions such as anger, so he cannot understand why John responds to the news happily.

10. **Spike said "Let's play a game! I bet I can defeat you"**

    If John responses yes then the system moves to the next state, the game state where John can play a smile game with Spike.

Case Study 4: Play with Poppy, User Win

1. **Poppy asks John to play a game.**

   The system moves to the next state, the game state where John can play a smile game with Poppy.

2. **Poppy asks John whether John wants to hear the rules or not.**

   The system asks the user whether they want to hear the rules or not. If they want to hear it, the system explains the rules; otherwise, the game is started.

3. **John says he wants to hear the rules.**

   The system asks the user whether they want to hear the rules or not. If they want to hear it, the system explains the rules; otherwise, the game is started.

4. **Poppy explains the game's rules briefly, and asks John if he is ready to play.**

   The system explains the rules and asks the user whether John is ready or not. If John is ready, the game then is started; otherwise, the system waits until John is ready.

5. **John says he is ready.**

   When John says that he is ready, the game is started.

6. **John attacks Poppy with raising his eyebrow continuously.**

   The system captures the user's facial expression. Currently, the user can attack the system with AU1 and AU27.

7. **Poppy responds "Good Move! Now is my turn"**

   Poppy responds to the user's facial expression.

8. **Poppy tries to counter John's attack with jokes, she/he says "I feel like such a homo sapiens right now."**

   The system has several movements and utterances to attack the user.

9. **Poppy reaches their first smile threshold and slightly smiles. She continues to attack John with raising her eyebrow.**

   If the game companion reaches their first smile threshold, the game companion will show a small smile. Poppy and Spike have a different smile threshold. Based on their personality, Spike has a slightly higher threshold than Poppy.

10. **John still able to hold his smile and attacks Poppy with raising his eyebrow and open his mouth.**

    The system captures the user's smile if the user reaches a maximum threshold of the smile, the user loses.

11. **Poppy cannot hold her smile; she reaches her second smile threshold and smiles.**

    If the game companion reaches their second smile threshold, the game companion will smile. Poppy and Spike have a different smile threshold. Based on their personality, Spike has a slightly higher threshold than Poppy.

12. **Poppy says "Good game, you win!"**

    If the user wins, the system will announce that the user wins.

Case Study 5 : Play with Spike, User Lose

1. **Spike said "Lets play a game! I bet I can defeat you"**

   If John responses yes then the system moves to the next state, the game state where John can play a smile game with Spike.

2. **Spike asks John whether John wants to hear the rules or not.**

   The system asks the user whether they want to hear the rules or not. If they want to hear it, the system explains the rules; otherwise, the game is started.

3. **John says he does not want to hear the rules.**

   The system asks the user whether they want to hear the rules or not. If they want to hear it, the system explains the rules; otherwise, the game is started.

4. **John attacks Spike with raising his eyebrow continuously.**

   The system captures the user's facial expression. Currently, the user can attack the system with AU1 and AU27.

5. **Spike responds "Meh! Now is my turn"**

   Spike responds to the user's facial expression.

6. **Spike tries to attack John with jokes, and he says "I feel like such a homo sapiens right now."**

   The system has several movements and utterances to attack the user.

7. **John still able to hold his smile and attacks Spike with raising his eyebrow and open his mouth.**

   The system captures the user's smile if the user reaches a maximum threshold of the smile, the user loses.

8. **Spike continues to attack John with raising his eyebrow.**

   The system has several movements and utterances to attack the user.

9. **John reaches his limit, and smiles.**

   The system captures the user's smile if the user reaches a maximum threshold of the smile, the user loses.

10. **Spike says "Hah! You Lose!"**

    If the user loses, the system will announce that the user loses.

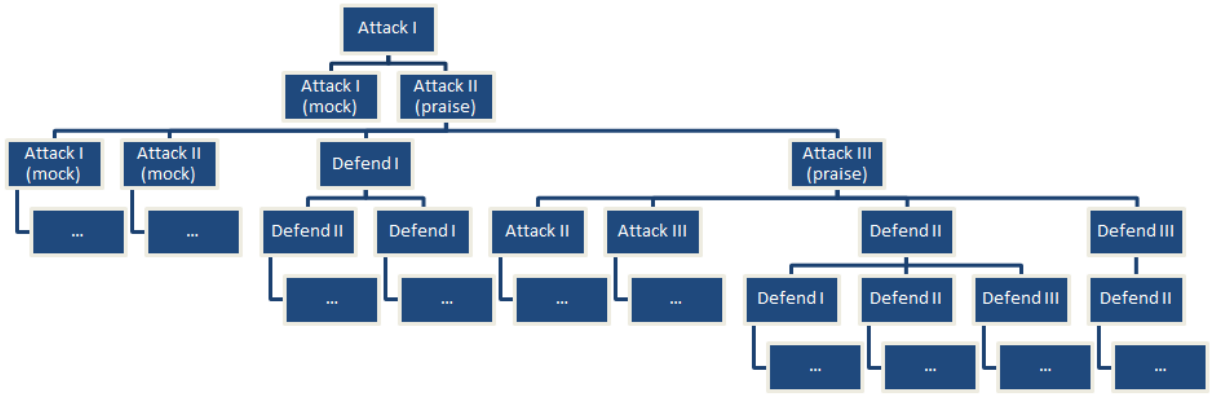


Figure 4.6: Game Levelling

4.3.3 Game Levelling

In the game, a level is one of the factors that makes the game is enjoyable. In The Smile Game, there are three levels of difficulty based on the game companion's urge to laugh (see Appendix 1). Hence, the game companion will show a set of behaviours that match to the ability of the player to make the game companion laugh. The closer the player to make the game companion laugh, the funnier (i.e., the harder) the behaviours that game companion will show. The game companion has two strategies in this game: attack and defend strategy. In an attack strategy mode, the game companion will show a set of behaviours that will make the player laugh by making funny or weird movements or jokes. While the defend strategy move will be activated when the game companion is about to laugh and try to mask their laugh behaviour conveyed through their facial expressions by using other movements (see Figure 4.3 and Figure 4.4). Figure 4.6 illustrates the levelling system in The Smile Game. If the player is making good progress with their attack, the game companion will reward them with praise such as "You are good. I almost laugh. It is my turn now". On the other hand, if the player's attack is not efficient enough to "tickle" the game companion, the game companion will encourage them with a word such as "Is that all you got? Come one, make me laugh".

To determine the level of the behaviours shown by the game companions is a subjective task. Hence, to rate the game companion's behaviours for the game levelling, we made 2 to 5 seconds videos with Poppy demonstrated 43 behaviours compiled from the video annotation

(see subsection 4.2.1 and Appendix 1). Subsequently, we recruited a total of forty-two participants via game forums (IGDA International, IGDA Indonesia and Gamedevid.org). The participants were from several cultural backgrounds, but most of them were from Indonesia. The participants were asked to rate with 5 Likert-scale on how likely the behaviour make them laugh/smile? Where 1 is unlike to make them smile, and 5 is given to the behaviour that will make them laugh out loud. Figure 4.7 demonstrates the average of the score for each behaviour. Behaviour 1 has the lower score with an average of 1.79, while Behaviour 9 rated the funniest behaviour showed by the game companion with an average score of 3.7. Moreover, there is no significant difference in score between the attack with jokes and facial expressions ($p = 0.246$). Attack behaviour with jokes does not necessarily get a good score as most likely the participants come from various culture and social background. This is only our assumption only based on literature in social science (Alden et al. 1993, Raskin 2012) as we do not have enough data to conclude. The other factor could be the participants thought the jokes were not appropriate to say to someone who we just met (Fine 1983). Moreover, behaviours that only using head movements seems to be rated to be in the lowest score category, while behaviours with facial expressions, in particular, the one involving muscles in the mouth area, were perceived to be the funniest behaviours compare to the other. Combination of head movements and facial expressions improve the chance to make the participants laugh/smile. Furthermore, we also added several additional behaviours for jokes and defend moves, for game levelling balancing. All the behaviours from the first to the last level are illustrated in section Appendix 1.

4.4 Behaviour Rules

We implemented an information states approach and rule-based classifier for the action selection method. In addition, we adopt a template-based method from The SEMAINE project (Schroder et al. 2012, ter Maat & Heylen 2011) to update the information states. The template describes the behaviours to be performed when a specific rule is fired. Figure 4.8 illustrates the rule-based classifier framework on our system. The Manager manages

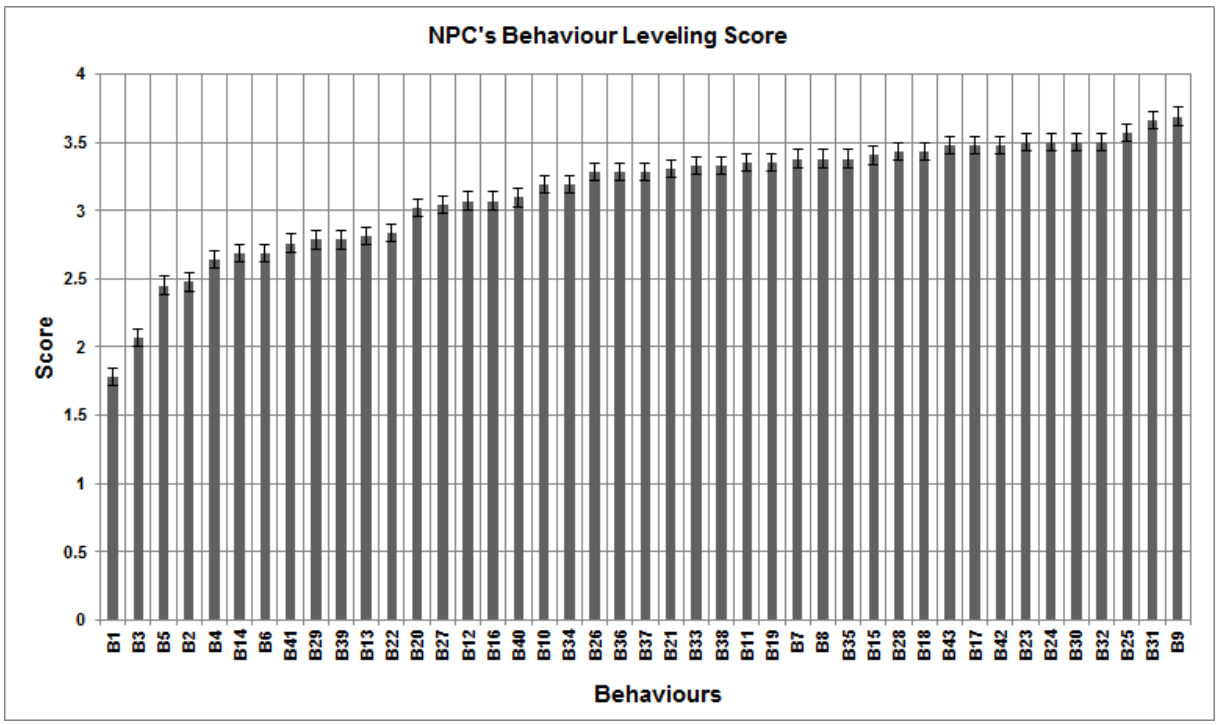


Figure 4.7: Agent's Behaviour Levelling Score

the inputs from external (i.e. player states), internal (i.e. game companion states) and indeterminate (i.e. dialogue and game states) stimuli and then it proposes an action to be executed based on the rules table. There are three main components in the rules classifier: dialogue states, turn-taking model and emotions, personality and social relationship models. The rules were derived from the datasets of interaction videos between human-human and human-agent, The SEMAINE Database Corpus (McKeown et al. 2012), as well as The Human-Human Interaction Data Collection, described in the previous Section.

There are four categories of Information States in this system: player, NPC, dialogue and game states (see Table 4.7). The player states contain information about the perceived player's behaviours from the Sensing and Interpreter Components such as their emotion, intention and when the player spoke last. The NPC States contain information about its belief about itself such as its emotion, personality, relationship values with the player, gender, date of birth and the probability of its need to speak. The game states contain information about the game itself such as started, paused, practice, ended, etc. Moreover, the game states store knowledge about whose turn it is in the game, the winner, current score, time elapsed and the game level. Additionally, game designers can append game states in accordance with

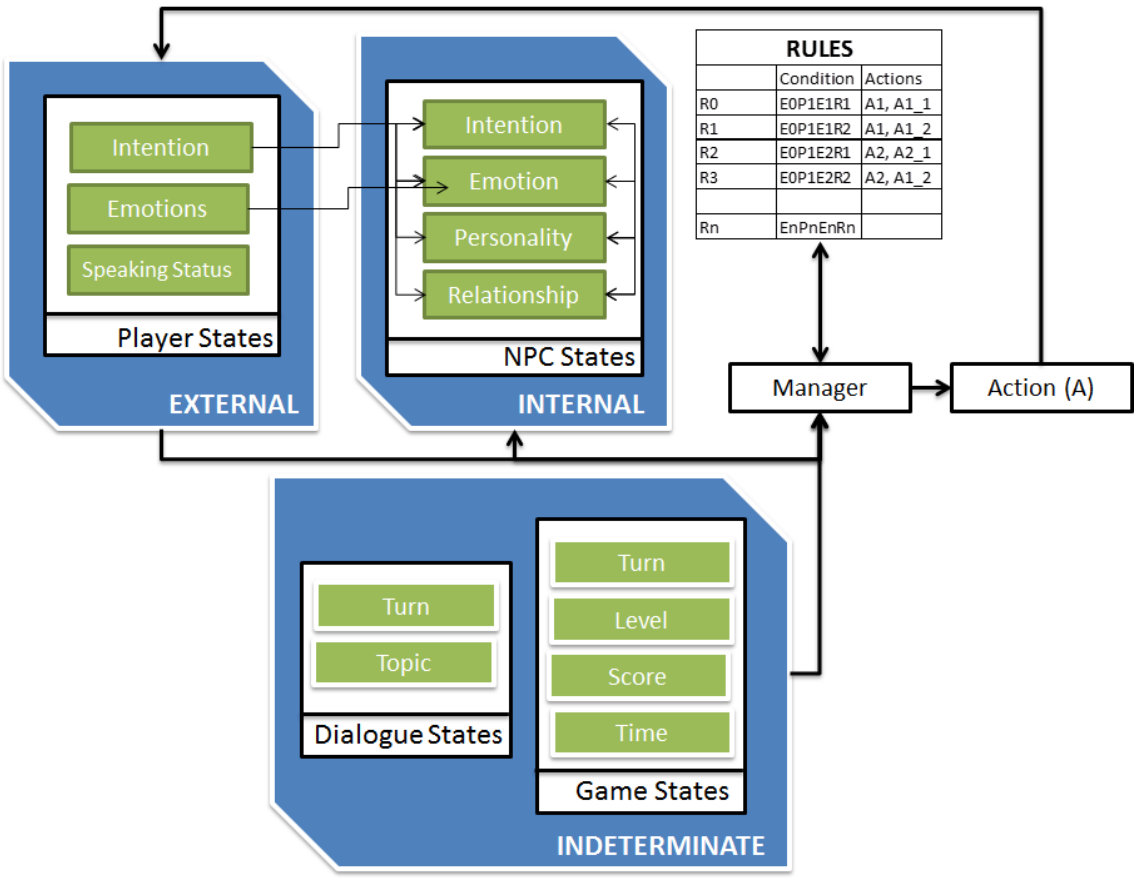


Figure 4.8: Rule Based Manager

the game design. Finally, the dialogue states consist of information about the interaction between the NPC and player, such as the conversation topic and whose turn now in the conversation.

The Interpreter Components of the Framework are responsible for interpreting the player's emotions and utterances. A keyword spotter and pattern matching methods are implemented in this work. A detected keyword or pattern will trigger a user intention thus update the dialogue and game states by using a template-based method from The Semaine project. As the states are changed, specific behaviours are triggered and performed through the embodiment system. For example, in Listing C.1, a keyword *hello* or *pattern hallo ** will trigger an update in the player states (*playerstates.intention*) and the dialogue states (*dialoguestates.topic*). The template will then decide which behaviours should be performed based on the behaviour file in Listing C.3 and C.4.

Table 4.7: Information States

Player States	
States	**Values**
playerstates.emotion.anger	Float:[-1, 0]
playerstates.emotion.disgust	Float:[-1, 0]
playerstates.emotion.fear	Float:[-1, 0]
playerstates.emotion.hapiness	Float:[0, 1]
playerstates.emotion.surprise	Float:[-1, 1]
playerstates.emotion.sadness	Float:[-1, 0]
playerstates.intention	String: greeting, goodbye, etc.
playerstates.isSpeaking	Boolean: true, false
playerstates.lastTimeSpeak	Long: 45844451
Dialogue States	
dialoguestate.topic	String: greeting, goodbye, game, etc.
dialoguestate.turn	Short [0, 1]
Game States	
gamestate	Integer: 1 for practice, etc.
gamestate.turn	Short [0, 1]: 1 for NPC
gamestate.winner	Short [0, 2]: 1 for agent win, 2 for draw
gamestate.score	Long: 5200
gamestate.time	Long: 5445646
gamestate.level	Integer: 1, 2, etc.
.....	....
NPC States	
npcstates.emotion.anger	Float:[-1, 0]
npcstates.emotion.disgust	Float:[-1, 0]
npcstates.emotion.fear	Float:[-1, 0]
npcstates.emotion.hapiness	Float:[0, 1]
npcstates.emotion.surprise	Float:[-1, 1]
npcstates.emotion.sadness	Float:[-1, 0]
npcstates.personality.O	Float:[-1, 1]
npcstates.personality.C	Float:[-1, 1]
npcstates.personality.E	Float:[-1, 1]
npcstates.personality.A	Float:[-1, 1]
npcstates.personality.N	Float:[-1, 1]
npcstates.relationship	Float:[-1, 1]
npcstates.affinity	Float:[-1, 1]
npcstates.familiarity	Float:[0, 1]
npcstates.gender	Short: [0, 2]
npcstates.DOB	Date
npcstates.DOB.date	Date
npcstates.DOB.month	Date
npcstates.DOB.year	Date
npcstates.needToSpeak	Float:[-1, 1]

4.4.1 Dialogue States and Vocabulary

The dialogue states are categorised into three simple categories: Greeting, Chit-Chat and Goodbye. In the beginning of the interaction, the game companion engages the conversation by greeting the player. The game companion prepares a preamble topic such as "evaluating the day" before entering the Chit-Chat state. In the Chit-Chat state, Both the game companion and player can start a particular topic. The interaction will keep going on until keywords or patterns related to saying goodbye are detected (e.g. **bye*, I have to go** or **see you**).

The game companion's data such as the dialogue, social relationships data, known faces databases, etc. are stored in XML based files. Listing C.2 shows the structure of the game companion's consciousness about itself (i.e. full name, date of birth, gender and personality). This is a static XML file and never supposed to be updated apart from initial design or for adjustment purposes (e.g. for game balancing). Listing C.3 exhibits the example of game companion's utterances vocabulary. The utterances are organised in categories, which are obtained from the data collections from the interaction between human-human and human-agent interaction described in Section 4.4. Each game companion has their style of utterances depending on their personality. In addition to the vocabulary of the game companion's utterances collected and manually crafted from the interaction videos described in this chapter and chapter 5, the agent has the recollections of their past story formulated in Listing C.4. The background story collections are collected from the past events that are significant to the game companion. Research indicates that virtual human who can recall and tell a background story from their past experiences, enhance their believability (Riedl & Stern 2006, Ho & Dautenhahn 2008, Li et al. 2014). In our system, a particular keyword or pattern detected from the player, depending on their social relationship level, will trigger the game companion to their past story and share it with the player. The game companion also has an attitude towards the story.

4.4.2 Response to Emotions, Personality and Social Relationships

In addition to the verbal responses, simple models of response to emotions, personality and social relationships were crafted for the game companions in addition to the normal responses. By default, each game companion has its basic answers, and we can augment the utterances with the player's and game companion's emotions, game companion's personality and their social relationships level using models described in the sub section3.3. For example: Suppose a default utterance to reply a greeting (category = *greeting*) from the player is *"Hello!"*, then the augmented version with a happiness emotion induced to the game companion (category = *greeting-posEmotion*) is *"Hello! It's a lovely day today isn't it?"* with happiness conveyed through the game companion's facial expressions. Moreover, if they have a positive relationships (category = *greeting-posRelationship*), the utterance will be *"Hello! $name!"* and if all the augmentation combined (category = *greeting-posEmotion-posRelationship*), the response will be *"Hello! $name! It's a lovely day today isn't it?"* or *"Hello! $name! It's good to see you"* with a happy face. The only problem with the model is that due to a high number the possibility of the variables combinations, it requires a lot of amount of time to spent to craft adequate rules for interaction.

4.4.3 Turn-Taking

The turn-taking rules manage the flow of the conversation between the game companion and player. The rules also decide when the game companion should start speaking or listening. There are three simple conditions in the rules: the player speaks, the game companion speaks, and no one speaks. When the player speaks, the relevant actions for the game companions are listening, back-channelling and interrupting. However, in this work, we do not include the interruption model for the game companion. If the player is interrupting the game companion when it is speaking, the game companion will stop the utterance and starts to listen to the player's utterance and update all the states. Finally, to keep the flow of the conversation as smooth as possible, the dialogue system should keep track of the conversation time. The player should not wait too long for a reply nor get it too soon

resulting in the overlapping to the player's utterance (unless it is intended). Maat suggested that the ideal waiting time if the player is not responding to the system, is 2 to 4 seconds (ter Maat & Heylen 2011). If the threshold time is reached, then the game companion will take the conversation floor and check whether the player is still there.

4.5 The Pilot Study: Experimental Setup

To explore the potential of the proposed framework and the computational models as well as to collect initial data for the game companions behaviours, we conducted a pilot study of *The Smile Game* in a controlled lab setting (see Table 3.5). We invited students and staff members of the University of Nottingham to play with both game companions. Each participant interacted with each of the game companions twice to evaluate the social relationships models as well as the Face Recognition Module. The game levelling system was not implemented in this study. Hence, the game only had 1 level only. However, Poppy was easier to defeat than Spike as the results of the design of the character's personality.

The first version of The Smile Game study was intended to be a pilot study to explore the potential of the models and framework as well as to assess the player's sentiment towards virtual game companions that are capable of conveying emotions and having personality. This study had six-teen adult player participants (8 Male; All Asian; AVG Age = 24.43; SD Age = 3.80; All Students) with a total of sixty-four interactions with the system. Each participant played two sessions of the game, and in each session, they played against both game characters (Poppy and Spike) while being video recorded from the player's profile position. After the game, they were asked to fill in a five Likert-scale questionnaire with a score of 1 indicates the lowest score, 3 shows the average score and 5 is the highest score (see B) to rate their perceived relationship with the game companion, the naturalness of the interaction during the conversation and the game and the game companion's response time. In addition, the participants were also asked to rate the game companion's personality. The players who were not familiar with the personality traits were explained and given the

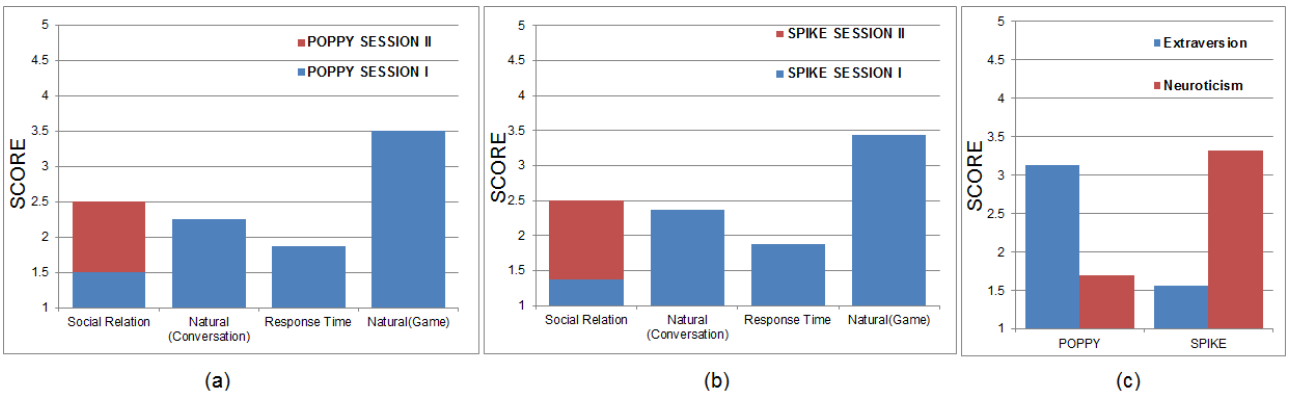


Figure 4.9: Evaluation Score of The First Version of The Smile Game. Quality of the game (a) and (b). Personality (c)

characteristic of the traits before filling out the questionnaire.

There were two phases in the game in this study: Conversation and Game phase. When the system detected a face, the game companion initiated a conversation with asking the player's name if that was the first time the player met the game companion or greeted the participants if the system successfully recognised the player. In this phase, the player was given the freedom to talk about whichever topic they like with the game companions. The game started when the system detected an utterance from the player with pattern *"*start*game*"*. This pattern triggered the system to start the Game phase. In this phase, the participants were only allowed to perform behaviours that only related to the game (i.e. a conversation about topics that related to other than the game was not possible in this phase). The game will stop when someone wins, or the system detected an utterance from the participants with pattern *"*stop*game*"* or *"*end*game*"*.

4.6 The Pilot Study: Results

4.6.1 Personality

Figure 4.9 (c) shows the score of the game companions' personality perceived by the players. Except for the rating of 'Social Relation', all scores were identical for session 1 and 2. The participants were asked to rate the game companions' personality subjectively with two

traits of Extraversion and Neuroticism. All the participants were explained briefly about the characteristics of each trait before they filled the questionnaire. Based on the result of the questionnaires, the personality scores of both characters are consistent with the personality that set to both characters. Poppy was perceived as a person with medium Extraversion and low Neuroticism traits, while Spike was perceived as a Neurotic person with low Extroversion trait. This indicates that the participants were able to differentiate the differences between the game companions' personality.



Figure 4.10: Player's Experiences when (a) distressed when the character could not understand their utterance, (b) enjoyed the game

4.6.2 Social Relationship

Figure 4.9 (a) and (b) shows the average scores for the social relationship for both Poppy and Spike. The participants were able to tell the changes in their social relationship with both characters in the second session of the game. The changes in the relationship are not very high but still significant because the participants only play with each character twice. A two-tailed t-test indicates that there was a significant differences between player's social relationships score in the first session compare to the second session for both Poppy ($t = -6.33, p < 0.001, SD = 0.63$) and Spike($t = -6.26, p < 0.001, SD = 0.72$). The player had the perception that the game companions recognised them in their second encounter.

The system was able to correctly recognise the players with Face Recognition Module during their second encounter with 87.5% and 93.75% success rate for Poppy and Spike respectively. However, the player somehow did not see how the relationships would change the quality of their interaction during conversation or playing the game, resulting in equal values for all other aspects of the questionnaire. Again, this due to the low amount of time spent with

the characters. This study provides insufficient data to analyse this. Hence, it would not be possible to draw a satisfactory conclusion about the social relationship factor.

4.6.3 Naturalness and Response Time

Figure 4.9 (a) and (b) shows the average scores for the naturalness of the interaction as well as the game companions' response time during the conversation and playing the game. The rating of the naturalness in the interaction during the conversation is quite low for both of the characters due to the long response time of the game companions. Upon further analysis of the problem, it turned out that this was caused by the low performance of the Speech Recognition Module. The hand-crafted rules for conversations turned out not sufficient to cope with the interactions during the study. Hence, further data collections for the conversations in the context of The Smile Game was collected throughout the next user studies in The Smile Game. From the video, it is clearly can be seen that most of the player had a difficulty at some point in the interactions during the conversation phase (see Figure 4.10). In contrast, participants gave quite high scores for the naturalness of the interaction rating during the game, probably at least partially because the majority of participants never used words/utterances to attack the game companions. The players start to enjoy the game when there were no words used in the interaction (see Figure 4.10). A two-tailed t-test indicates that the naturalness score by the players in the game was significantly higher than the one during the conversation phase for both Poppy ($t = 7.32, p < 0.001, SD = 0.68$) and Spike ($t = 4.70, p < 0.001, SD = 0.96$) in both sessions. However, there were no significant differences between Poppy's and Spike's evaluation scores except for their personality score rated by the players.

4.7 Conclusion

This chapter presents the methods to build game companions in a simple game called The Smile Game by using the proposed framework and computational models. The purpose of the

Smile Game was to evaluate the proposed framework and computational models in a game environment. We postulate that the proposed framework and computational models will provide a unique characteristic to each game companion through the computational models as well as the ability to perceive social relationship they have with the players. Those eventually will provide more natural game companions. In this chapter, we demonstrate how to collect and craft the interaction rules for the virtual NPCs in the game environment to support the creation of affective loop experiences? **RQ4**. Two datasets were used to design the game companions. The first dataset contained the interaction rules when people were playing The Smile Game. The data were collected from the human-human interaction playing The Smile Game. The second dataset contains the conversation rules, and the data were gathered from The SEMAINE Database McKeown et al. (2012).

Moreover, this chapter also illustrates how to establish natural and believable interaction between a player and game companion to create affective loop experience that is informed by the process of encoding and decoding of social signals exchanged between the player and game companion? **RQ1**. The results show that the participants were able to tell the differences in the game companions' characteristics. Furthermore, the affective loop experiences between the participants and the game companions creating a natural and believable social relationship between them (**RQ2**). The participants were able to tell the changes in their social relationship immediately in the second session of the game ($t = -6.33, p < 0.001, SD = 0.63$ for Poppy and $t = -6.26, p < 0.001, SD = 0.72$ for Spike) (**RQ3**). Although the player somehow did not see the relationships would change the quality of their interaction. This most likely due to the low amount of time spent with both characters. Hence, this study provides insufficient data to analyse and draw an adequate conclusion about the social relationship factor. The response time and naturalness score in the conversation phase were rated relatively low in both characters. This due to two major reasons. The first was because the Speech Recognition Module was too sensitive with the background noise and the quality of the built-in microphone is insufficient to capture the player's speech. The second reason was due to the hand-crafted rules for conversation turned out insufficient to cope with the interactions. This was reflected in the response time and naturalness score rated

by the players in the interaction that did not required the Speech part (i.e. game part) got higher score than in the conversation phase ($t = 7.32, p < 0.001, SD = 0.68$ for Poppy and $t = 4.70, p < 0.001, SD = 0.96$ for Spike).

Chapter 5

Human - Virtual Game Companions Interactions in the Wild

5.1 Introduction

After an improvement to the Smile Game was applied based on the evaluation of the pilot study, the system was deployed outside of the lab environment. A series of studies were conducted in the semi-public and public environment. These studies aim to explore human and virtual human interaction in the wild in a game setting. As stated in Section 2.10, there are not many virtual human application deployed in a real-world setting as there are too many noises and have the unlimited number of possibilities of the interaction between them and the players. Hence, by having a virtual human that able to interact with people in the wild affectively, we as researchers can learn something from the interaction. In the series of studies, we aim to explore: Human's behaviours when they interacting with the virtual human in the semi-public and public environment; The human's perception (e.g. personality and emotions) of the virtual human (i.e. Poppy and Spike); What are the favourite topics and attacks that player (human) shows; Human's emotions (i.e. valence and arousal) during the interactions. We hope that by learning something from those aspects could lead us to provide some insight and recommendation to researchers who are interested in deploying

virtual human in the wild in a similar setting with ours. For example, The favourite topics could be used as an indicator of what would usually people ask or say to the virtual human in a public space setting. Hence, we can prepare knowledge of those topics to the VH.

5.2 Wizard of Oz Study in Semi-Public Space

In the semi-public space, a semi-automatic system using Wizard of Oz method and a fully automatic system was conducted to identify the advantages as well as the disadvantages of both systems. As suggested in the pilot study, The Automatic Speech Recognition Module was not sufficient to adequately handle the conversation. Hence, in the semi-automatic system, *The Wizard*, performed by the researcher, had control of The Automatic Speech Recognition Module in The Sensing Components, The Verbal Interpreter Module in The Interpreter Components and all modules in The Behaviour Components. While the other components, remain to be running automatically. *The Wizard* was able to determine which parts *The Wizard* would like to control during the interaction. During the conversation, *The Wizard* was able to provide feedback to the player by choosing a specific behaviour manually or simply type the category (i.e. the dialogue acts) and let the system determines a behaviour automatically based on *The Wizard's* input in addition to the other inputs from The Sensing Components and the NPC's internal states (e.g. emotions, personality, and their relationships).

The study was intended to evaluate the system and computational models further as well to gather more behaviours data for the game companions. In this study, the system was installed at one of our school's *semi-public* spaces, where it is only accessible to a fixed group of people. Any person (age of 18 and over) who passed by the area was invited to interact with the system by the game companions. Hence, it established a natural space for interaction with the virtual game companions, allowing for serendipitous encounters.

In this study, the interaction phases were extended to four from two in the pilot study. This allows the flow of the conversation become more natural compared to having only two

phases. The phases are Opening, Chit-Chat, Game, and the Closing phase. The Opening Phase is a phase where the participant initiated their interaction with the game companion. In this phase, the game companion greeted the participant and asked the participant's name if the game companion had not met the participant before. Afterwards, the game companion initiated a small chit-chat using "evaluating the day", weather or news dialogue (Chit-Chat Phase). The next phase was the Game Phase where the participant and the game companion played the game trying to make each other laughed by making weird facial expressions, funny body movements and jokes. Finally, the game companion evaluated the game and said goodbye after they finished the game (Closing Phase). To avoid any ethical issues during the interaction, a video and audio recording was made using a clearly visible camera and shotgun microphone.

As the results of the pilot study indicated a problem in the robustness of the ASR Module, we implement a Wizard of Oz method to replace the ASR Module in this study. We then compare the results with a fully automatic system with the same settings but different participants. The participants were not aware that the game companions was a semi-automatic system (i.e. some components were controlled by an operator).

5.2.1 Experimental Setup

Two different methods were applied to this study. In the first method, the participants (age of 18 or over) were personally invited by the researcher to interact with the system. While, in the second method, we run the game system at random times during working days for approximately 1.5 hours per day in 2 weeks time frame, and let any person (aged 18 or over) who was passing by, interacted with the system. The participants were allowed to interact with the virtual game companions as long as they want and as many time as they like. The virtual game companions will forge a relationship link to the participant, depending on the quantity and quality of the interaction with the participant.

In this study, an operator called *The Wizard* controlled several system functions manually while a player interacts with the system, believing that they are interacting with a fully func-

tional application system (Green & Wei-Haas 1985*b*). The Wizard replaced The Automatic Speech Recognition modules in The Sensing Components, The Verbal Interpreter modules in The Interpreter Components, and all modules in The Behaviour Components. While the other components, remain to be running automatically. In *The Wizard's* screen (See Figure 5.2 no. 1), *The Wizard* observes the participant's face through a web camera and hears their voice through a microphone. In addition, *The Wizard* monitors participant's internal states through a graphical representation chart(as shown in Figure 5.2). *The Wizard* then interprets the participant's behaviour and creates a decision to reply to the participant using command line by choosing actions provided in the screen.

During any phases, *The Wizard* can either choose a specific action or just the category and let the system determines a specific action based on the participant's and virtual game companion's internal states. For example: In the greeting phase, *The Wizard* can either choose a specific greeting utterance or type "greeting". The system will then choose a greeting utterance for *The Wizard* to greets the participant. Another example: In the game phase, *The Wizard* can specify what kind of facial expressions or jokes to perform or type "attack" or "defend" to perform an attack or defend behaviour based on the virtual game companion's urge to smile. The closer the participant is, to make the virtual game companion laugh, the more aggressive the virtual game companion becomes.

There were nine distinct adult participants (8 males) with 12 interactions with the virtual game companions (10 with Poppy, 2 with Spike) from a period of time of 25 February to 10 March 5015. The total duration of the interactions was 30 minutes and 48 seconds ($MIN = 36s; MAX = 333s; AVG = 153.42s; SD = 100.35s$). Six participants were invited to interact with the virtual game companions, while the others spontaneously played with the virtual game companions. In addition, there were ten interactions where the participants interacted for the first time, with the virtual game companions.

5.2.2 Results

In Wizard of Oz Study in Semi-Public Space, not all the players finished their game with the game companions. There are only 56 % of the players finished playing the game. This due to two major factors, either the game was too hard, or the players did not understand how to play the game. The players who played with Spike believed that Spike is impossible to defeat, as he appeared to be someone who is serious, mean and hardly smiles. All players who had the game with Spike stopped playing before the game ended because they found it too hard to make him laugh and ultimately got bored. Furthermore, most of the players were not sure how to play the game as they were new to this type of game and had never played the game before. They were not sure what kind of expression to perform to attack the game companions and win the game. As a result, two of the players decided to stop playing in the middle of the game.

From 4 phases, 11 participants went through the opening phase. One of the participants only looking at the agent without saying anything despite the game companion already initiated the conversation. Nine participants went through the chit-chat phase while the others broke down the interaction because they did not want to tell their name. All nine participants then went through the game phase, but only seven stayed until the closing phases with 2 of them not finishing the game. The cause of the low completion rate was most of the participants were confused about the game rules, when the game started and how the game turns work. The participants were not sure what kind of movements were allowed in the game, some of the participants asked the game companions repeatedly about the game rules (There were a total of 15 conversations about the rules with a total average of 23 seconds duration needed to discuss the rules). The game turn was also a problem for the participant. They were not sure when to start the game, some participants started to attack the game companion even though the game had not started yet. Moreover, the recording showed that some of the participants were not sure whose turn it was at that moment.

Among the 5 participants who finished the game, there was only one participant who won the game playing with Poppy (winning rate 20%). Others could not help but laugh either because

of the backfire of the movements they did or the successful attacks performed by the virtual game companions. Mostly, the game companion won due to their funny facial expressions. 2 participants laughed immediately after they performed funny facial expressions. This indicates, they were laughing because they thought they were doing some funny and stupid things. In four out of twelve game sessions, the participants also mimicked the virtual game companion's facial expressions as their strategy to win the game when they were running out of ideas which expressions they should show to attack the game companions. In contrast with Spike, Poppy was perceived as someone who is nice, likeable and friendly game companion (see Figure 5.5 and 5.4). In this study, all players who interacted with Poppy exhibited positive emotions during the interactions and displayed negative emotions in some point of the interaction when playing with Spike (see Figure 5.5 and 5.7). However, interestingly, most of the participants prefer Spike's voice because sometimes Poppy's words were hard to understand.

Topic Analysis

Table 5.1 illustrates the conversation topics in WoZ Study between players and the game companions during the interaction from the interaction videos (1920 x 1080 and 29 fps) annotations. Topic about the game appeared to be the most popular among the other topics followed by topic about the game companion (greetings was not included as greeting and goodbye are most likely appeared in every interaction). The players had a different attitude towards the game and the game companions as mentioned above. There were 11 conversations which the players expressed their opinions towards the other game companion. Six of the conversations were negative opinions expressed towards Spike, four utterances were positive opinions expressed towards Poppy, and a negative opinion was expressed toward Poppy related to her utterances were sometimes hard to understand.

Table 5.1: Topics in The WoZ Study at Semi-public

Annotation	Occurrences	Total AVG Time(s)
About Agent	**25**	**64.82**
ask-wh-question-aboutagent	24	63.17
statement-opinion-aboutagent	1	1.65
About Game	**92**	**177.35**
statement-nonopinion-aboutgame	30	44.69
ask-wh-question-aboutgamerules	15	23.21
answer-yes-no-question-readytoplay	17	25.76
ask-wh-question-aboutgame	2	4.97
ask-yes-no-question-playgame	10	25.84
ask-yes-no-question-readytoplay	11	13.69
statement-nonopinion-aboutgamerules	7	39.19
About Player	**2**	**4.23**
statement-opinion-aboutplayer	2	4.23
About News	**21**	**53.19**
statement-opinion-aboutnews	1	1.77
ask-wh-question-aboutnews	11	22.46
answer-wh-question-aboutnews	9	28.96
About Other Agent	**11**	**11.94**
statement-opinion-aboutotheragent	11	11.94
Clarification	**19**	**18.81**
ask-clarification	11	12.58
ask-to-repeat	5	4.40
apology	1	1.83
About Weather	**3**	**4.97**
ask-clarification	2	3.20
ask-to-repeat	1	1.77

Table 5.1: Topics in The WoZ Study at Semi-public

Annotation	Occurrences	Total AVG Time(s)
About Movies	**8**	**3.84**
statement-opinion-aboutmovies	8	3.84
Greetings	**36**	**48.30**
greeting	36	48.30
Goodbye	**12**	**25.18**
goodbye	12	25.18
Error	**6**	**20.69**
statement-nonopinion-confusion	2	2.44
statement-nonopinion-wizarderror	4	18.25
About Life	**12**	**21.51**
statement-opinion-aboutlife	12	21.51

There were 181 game companion's and 146 player's turns with an average duration of 38.1 s and 27.2 s of interactions respectively (see Table 5.2). In the turn taking, sometimes there were long delays on average of 10.7 s of waiting time per turn on the game companion's responses due to the time needed for *The Wizard* to manually type the dialogue acts for the input. This caused some participants to repeat their previous utterance to make sure that the game companions were still listening.

Table 5.2: Turn Taking in WoZ Study

Annotation	Occurrences	AVG Time (s)	Latency (s)	AVG Latency (s)
Game Companion	181	38.1	38.4	3.2
Player	146	27.2	128.1	10.7

Attack Analysis

Facial expressions were the most popular attack to perform as a strategy to make the game companions laugh followed by jokes as the second choice of the strategy (see Table 5.9). A



Figure 5.1: Movements Type

total of 34 attacks executed by the players to the game companions by pulling their facial muscles, with a total of average time of 30 seconds (average of 0.88 seconds/attack). In the second place, a total of 26 jokes were performed in a total of average time of 29 seconds (1.12 s/attack) as the strategy to win the game. Also, some players performed a combination of body movements (e.g. shoulder or hand movements) with their facial expressions. Figure 5.1 illustrates all the attacks type performed in the Smile Game. The attacks, which were using other modalities than facial expressions, were not automatically perceived by the system. Moreover, the system only perceived three Action Units (AUs) (i.e. 1, 12 and 27) and its combination. AU 1 (Inner brow raiser) was sometimes detected as a false positive, AU 12 and 27 were fairly decent perceived by the system. Finally, although the game companions were able to recognise correctly the players who have interacted with them before, in this study, there was no sufficient empirical evidence to evaluate the social relationship model.

5.3 The Fully Automatic System Study in Semi-Public Space

To get more data for analysis if there are any differences between WoZ and Fully Automatic System in a Semi-Public Space, we deploy a fully automatic system of both Poppy and Spike version at the exact same location on one of our school's *semi-public* spaces, where it is only accessible to a fixed group of people. Similar settings with the WoZ study were applied to this study (e.g. the interaction phase, and the conversation strategy). Moreover, the participants did not aware that the game companions was a semi-automatic system (i.e. some components were controlled by an operator) or run fully automatic.

5.3.1 Experimental Setup

The method of this study is similar to the first study (The Wizard of Oz Study). The only difference in this study was that the system was running fully automatic. In this study, we run the game system at random times during working days for approximately 2 hours per day in two weeks time frame. The game companion would initiate a conversation if the system detected a face that is staying in the detection area long enough. A tape was placed on the floor to mark the interaction area. Hence, the participants were aware of the game companion's visual coverage. The game companion will try to maintain the flow of the conversation by using behaviour rules described in the section 4.4 until a specific keyword was detected. Similarly to the pilot study, if the system detected an utterance spoken by the player with a pattern like *"*start*game*"*, the system will start the game. In addition, the game companion will ask the participant to play a round of game if the time spent in the Chit-Chat phase was too long. This was designed to avoid a deadlock where the system failed to recognise the *"*start*game*"* keyword or pattern and the player could never enter the Game phase. The game ends when reached one of these conditions: a person wins, *"*stop*game*"* or *"*end*game*"* pattern detected or the time spent in the Game phase was too long. Again, this was designed to prevent a condition where the player could never end

the Game phase because of some circumstances such as no one wins, and the player starts to losing their enthusiasm to continue the game, or the system failed to detect the magic words to stop the game.

Both Wizard of Oz Study and Fully Automatic System Studies in *Semi-Public Space* were video and audio recorded from the perspective of both the player and the NPC to be further annotated. Furthermore, all the interaction videos were annotated using a multi-modal interaction annotation tool, ELAN (Brugman et al. 2004) and analysed for the evaluation. The videos were annotated for the interaction turn, player's engagement, game strategy, movement type, topics, winners and player's and NPC's smiles. The conversation topics were annotated to understand what participants usually talk about while interacting with the virtual game companions in this scenario. Figure 5.2 illustrates the interaction video annotated with ELAN. No 1 shows *The Wizard's*/Debug screen, No 2 shows the player's screen, and No 3 shows the annotation block. In addition, some of the participants from the semi-public space studies were invited to a short interview to discuss their interaction with the game companions.

Fully automatic system study in semi-public space had a similar setting with the Wizard of Oz Study in Semi-Public Space. Instead of controlled by *The Wizard*, the system run fully automatic in this study. The system was also installed at our lab meeting room bay (Room B07) for two weeks (from 13 - 20 April 2015). In this study, there were 12 distinct adult participants (8 males) with a total of 12 interactions with Poppy. The fully automatic system spent a total of 19 minutes and 31 seconds ($MIN = 3s; MAX = 351s; AVG = 97s; SD = 101.47s$) to interact with players. All participants were spontaneously played with the game companion without knowing what they should/can do with the game companion, except for one participant who was invited to play with the game companion. In addition, there were ten interactions where the participants interacted for the first time, with the game companion.

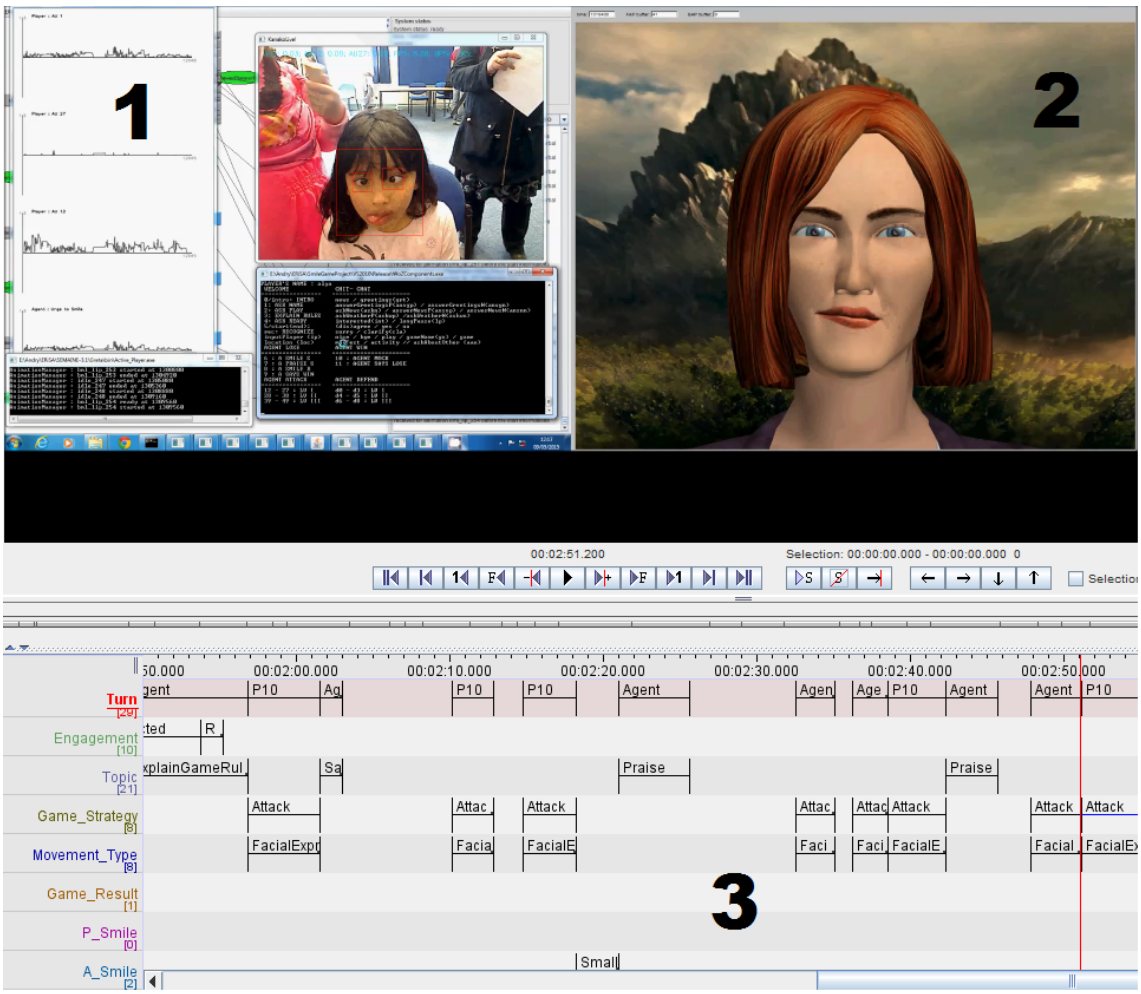


Figure 5.2: The Smile Game User Study

5.3.2 Results

Fully automatic system study in semi-public space, only 2 participants who were successfully played with Poppy, and only 1 finished the game(winning rate 50 %). Ten out of twelve participants were not sure what to do, and how to play the game while the game companion failed to interpret the user's speech due to the speech recognition technology. From the all of the four phases, nine participants successfully went through the opening phase. Two players were only looking at the game companion without a single word despite the game companion already initiated the conversation and ultimately left the game companion behind. The system did not recognise a player due to the lighting condition and the position of the participant that quite far away from the system (i.e. the Face Recognition Module could

not locate the player's face). From the other nine participants, only 2 participants who were successfully played with Poppy. 33.333% of participants were not sure what to do with the virtual human, how can they interact and play a game with the virtual human and also what is the objective of the game (16.667%) as there were no clear instructions provided in the study setting. Resulting most of the participants were not sure what to do, and how to play the game while the game companion failed to interpret the player's speech due to the speech recognition technology.

Similarly, with the WoZ user study, a participant was confused about the game rules and the game turn system. They were not sure whose turn it was during the game. This problem can be easily fixed with some adjustment in the virtual human's utterances to clarify whose turn it is now (e.g. simple word such as "I almost laugh, it is my turn now!" or "come on! it is your turn now!"). In addition, the participants were not sure what kind of movements that are allowed in the game and asked Poppy what kind of movements that are allowed to make her laugh. Eventually, the participant left the game companion because the participant was bored. There was one participant that played with Poppy have repeatedly interacted with the system. Hence the participant knows how to play with Poppy and already familiar with the rules more than the others.

Topic Analysis

Table 5.3 demonstrates the conversation topics in automatic system study between players and the game companions during the interaction from the interaction videos (1920 x 1080 and 29 fps) annotations. Similar with the WoZ study, a topic about the game itself was the most favourite topic in this study, with a total of 51 occurrences and a total of average time of 116.44 s was spent on talking about this topic (2.28 s/turn). The second favourite topic was news topic with 43 occurrences in a total of average time of 70.29 s (1.64 s/turn) was spent in this topic (greetings was not included as greeting and goodbye are most likely appeared in every interaction). There were 45 conversations with a total of average time of 40.10 s to spent in the clarification topic as the results of the Automatic Speech Recognition module

was failed to recognise the player's utterances and asked the player to repeat or clarify their previous utterances. In addition, as the results of the broken conversation between them, the player also asked the game companion to clarify their previous utterance.

Table 5.3: Topics in The Auto System at Semi-public

Annotation	Occurrences	Total AVG Time(s)
About Agent	**2**	**4.46**
ask-wh-question-aboutagent	2	4.46
About Game	**51**	**116.44**
ask-wh-question-aboutgame	6	9.43
ask-wh-question-aboutgamerules	9	13.48
statement-opinion-aboutgame	1	3.88
statement-nonopinion-aboutgamerules	5	56.68
ask-yes-no-question-playagain	2	2.36
ask-yes-no-question-playgame	17	16.70
ask-yes-no-question-ready-toplay	9	10.01
statement-nonopinion-aboutgamerules	2	3.90
About News	**43**	**70.29**
ask-wh-question-aboutnews	14	31.36
answer-wh-question-aboutnews	11	12.67
statement-opinion-aboutnews	18	26.25
About Presence	**12**	**21.28**
ask-yes-no-question-agentpresence	12	21.28
Bored	**1**	**3.14**
statement-opinion-expressboredom	1	3.14
Clarification	**45**	**40.10**
ask-clarification	22	20.81
give-clarification	15	14.62
ask-to-repeat	8	4.67

Table 5.3: Topics in The Auto System at Semi-public

Annotation	Occurrences	Total AVG Time(s)
Greetings	**34**	**81.52**
greeting	34	81.52
Goodbye	**2**	**3.90**
goodbye	2	3.90
Error	**9**	**31.74**
statement-nonopinion-confusion	9	31.74

There were 145 game companion's and 107 participant's turns with total average duration of 35.4 s and 14.1 s interactions respectively (see Fig. 5.4). Compared to the WoZ, the delays in the turn-taking were faster in this study. However, the turn-taking still requires the average of 4.5 s of waiting time per turn on the game companion's responses due to speech recognition delays and failure. This caused some participants had to repeat their previous dialogue to the game companion to make sure that the game companion is still listening or asking whether the game companion is still there and not broken.

Table 5.4: Turn Taking in Fully Automatic Study

Annotation	Occurrences	AVG Time (s)	Latency (s)	AVG Latency (s)
Game Companion	145	35.4	16.2	1.6
Player	107	14.1	40.8	4.5

Attack Analysis

Similar to the previous study, facial expressions and jokes seem to be the most popular strategy to attack the game companions (see Table 5.9). However, in this study, the attacks were executed relatively faster than the previous study. A total of 26 attacks executed by the players to the game companions by pulling their facial muscles, with total of average time of only in 6 seconds (average of 0.23 s/attack) and a total of 9 jokes were performed in a total of average time of 3 seconds (0.33 s/attack). The turn taking in the game (not in

the conversation) was turned out too fast. Also, some players also performed a combination of body movements (e.g. shoulder or hand movements) or head movements with their facial expressions (see 5.1). Finally, the system was 100 % successfully recognised the returning player. However, there was also no adequate data to evaluate the social relationship model in this study.

5.4 Enhancing The Smile Game

Based on the previous evaluation results, we improved the Smile Game by adding additional rules to both Poppy and Spike (Smile Game V.3 or SML V.3). The interaction rules were gathered and annotated from the previous interaction with Poppy and Spike in the Pilot, WoZ, and Fully Automatic System. Moreover, the Game Design was also enhanced (SMLGD V.2) to create engagement and immersion effect to the character (i.e. Poppy and Spike). The enhanced Smile Game was evaluated in our university public events (The Mayfest and The Open Days).

5.4.1 Additional Behaviours Rules

We annotated the interaction videos from the previous *The Smile Game* Studies (i.e. pilot, woz and automatic) with Dialogue Acts to see what would the players talk about with a virtual game companion in a game setting, specifically in *The Smile Game*. Hence, we append more data to the game companion behaviours. Table 5.5 demonstrates all the Dialogue Acts annotated from the player's utterances during the interaction. Nine-teen Dialogue Acts were extracted from the player's conversation transcripts with a total of 871 occurrences.

During the interactions, some of the players (67 utterances) were asking the game companions about general knowledge such as "Who is the president of The United States of America?" or "Tell me where is Jakarta?". The game companions could tell the player that they do not have the knowledge of the subject. However, to add more data to the game com-

Table 5.5: Player's Dialogue Acts in The Smile Game

Player's Dialogue Acts	**Count**
agreement	16
ask-wh-question-aboutagent	98
ask-wh-question-aboutgame	73
ask-wh-question-aboutgamerules	44
ask-wh-question-aboutgeneralknowledge	67
ask-wh-question-aboutmovie	5
ask-wh-question-aboutnews	22
ask-wh-question-aboutotheragent	26
ask-yes-no-question	68
disagreement	53
goodbye	97
greeting	112
repeat-phrase	21
statement-nonopinion-aboutplace-positive	12
statement-opinion-aboutagent	23
statement-opinion-aboutagent-negative	46
statement-opinion-aboutmovie	36
statement-opinion-aboutnews	36
statement-opinion-aboutwork-negative	16

panions knowledge by appending existed utterances data about general knowledge from the Annotated ALICE AIML[1] (Wallace 2003) to the system's responses database. The AIML files were chosen on account of that the responses provided in the files support keyword spotting and pattern matching method. Moreover, we also added several background stories for the game companions that are related to the game. For example: where or how did the game companion learn how to play the game or who is the best player, according to the game companion, in this game.

5.4.2 The Game Design

Schell argues that there are four main ingredients to craft a good game: mechanics, story, aesthetics and technology (Schell 2014). Based on the evaluation results from the previous study, we extended the game elements with simple mechanics for the Smile Game, the story about the characters and the game and improved the aesthetics elements.

[1]http://www.alicebot.org/aiml/aaa/

Mechanics

Game level influences the difficulty of the game, timer and the game companion's personality and emotion at time t (see Section 4.5). In this study, a stage system was introduced to replace the levelling system in the previous studies. With this system, the player was able to play with the game companions in several stages. The number of stages was limited to how fast the player make the game companions laugh. Moreover, a trial level was accessible to the new players in this study. In a trial level, the game companion explains the rules of the game, performed a funny facial expression as an example and asked the player to play 30 seconds of trial round. In the early stages, the game companions will only perform actions that belong to a level I category (i.e. Attack I of Defend I, see Section 4.3.3) and gradually increased accordingly with the stage.

Timer mechanic in the game (in seconds) is illustrated in the equation 5.1. Current game time set by an initial time set to the game and bonus time from the previous stage. The bonus time is derived from the remaining time from the previous stage and the game companion's intention to play determined by their emotion and social relationship towards the player at time t (see equation 5.2 and 5.3). All the mechanics are controlled in the Game Play Modules form the Game Components in the framework.

$$gtime = initial_gtime + bonus_gtime - 10 \tag{5.1}$$

where,

$$bonus_gtime = rtime + \mathcal{X}_{i,j,t} \tag{5.2}$$

and

$$\mathcal{X}_{i,j,t} = (max(avg(\mathcal{I}_{v+,i,j,t}), avg(\mathcal{I}_{v-,i,j,t})) + \mathcal{R}_{i,j,t}) * 10 \tag{5.3}$$

Finally, to motivate the player in playing the game, a scoring system was introduced in this game. The game score is calculated from the bonus game time, bonus action and how many times the player almost laugh. A bonus is given to the player every time they perform a movement with their facial expressions. This, however, is limited to the system's capability to detects player's AUs.

$$gscore = (bonus_gtime * 100) + sum(bonus_action) - (sum(AU12_p) * 50) \tag{5.4}$$

Story

A good story in a game will increase the player's immersion into the game. Player naturally emotionally involves to a game character who has a story behind the character. The game does not require complex stories to make it good. However, a game needs to shapes the game character, gameplay and the game world. In this game, stories are categorised into three groups: general knowledge of the game companion's (e.g. when and where they were born, their name and their relationship with the other game companions), background stories (where did they learn to play *The Smile Game* and what makes them love to play *The Smile Game*) and current situation (e.g. where are they now and what are they doing now). These stories are mostly reflected in the game companions utterances and are stored as the behaviour rules of the Behaviour Components. For example, in this game, we designed a character called "Poppy Williams" who was born at Salisbury on 4 May 1952. She has a colleague called "Spike Semaine" who is working in the same division as her. Her mom taught her how to play *The Smile Game* in her childhood. She likes to play it with her friends to release the stress. In her opinion, the game is simple, and she can play it anywhere, any time. On the day of the study, both characters are in their business trip to Scotland, and they use this opportunity to make a trip to Ben Lawers Mountain as both of them like hiking.



Figure 5.3: First (Left) and Second (Right) Version of Background

Aesthetics

To support the story, a 3D image of mountains scenery was set to the back of the game companions to provide an illusion that they are currently in the mountains area. A bird sound was also playing in the background during the conversation to build the ambience. Once the game is started, cheerful music is playing to build a mood. The tempo of the music is increasing dynamically based on the how close the player make the game companion laugh to create the tension. All the aesthetics functionality was regulated in the Game Play Module, the Game States and Game Rules Memories in the Game Components in the framework. Figure 5.3 illustrates the aesthetic difference between the previous and current game.

5.5 The Smile Game in Public Space: The Mayfest

Based on the evaluations of the pilot studies, we deployed an updated system and evaluated it during two *public* events at The University of Nottingham: Mayfest 2015 and OpenDays 2015. This section describes the study and evaluation of the system deployed in The UoN Mayfest Event 2015. The UoN Mayfest is an annual event open for the public to introduce what kind of activities we are doing in the university. The Mayfest's visitors are mostly children and their parents. In this study, any person of any age who showed an interest was invited to play with the characters by the attending researcher. All interactions were video and audio recorded. To adhere with the research ethics procedure, a consent from the parents or guardians was required if the participants were minors (aged 17 and below).

Due to the very noisy environment in the public space, it was not possible for the system to run fully automatic. Hence, we implemented a Semi-Automatic Wizard of Oz in this study. In contrast to the first study, in this study, The Wizard only controlled Automatic Speech Recognition modules in Sensing Components and part of Interpreter modules in Interpreter Components, while the other components remained automatic.

5.5.1 Experimental Setup

In the public event, The MayFest 2015, the system was running only for a day for approximately six hours in total. Any person of any age who showed an interest was invited to play with the characters by the attending researcher. During the interaction, a camera from the game companion's perspective was recording the interaction and the screen showing the game companion. This study aimed to evaluate the system and the second version of computational models. However, due to the nature of the interaction, it was not possible to evaluate the model of social relationship as the players are new to the system. By the nature of the environment, it also was not possible to evaluate the Automatic Speech Recognition and Verbal Interpreter Modules in this study. As it was impractical to have a fully automatic system in the public area due to the noise background, once again the Wizard of Oz method was implemented. In this semi-automatic system, *The Wizard* was only required to input a dialogue act or a specific keyword shown in the screen to the system, and it would automatically perform a behaviour as a feedback to the player during the conversation.

In total there were fifty-three distinct players (32 males, three teenagers, 34 children) with 58 interactions, played with both characters (50 with Poppy, and 8 with Spike) in both *public* events. The total duration of interactions was 2 hours 53 minutes and 2 seconds ($MIN = 35s; MAX = 559s; AVG = 169.1s; SD = 91s$). All players spontaneously played with the ECAs, except two people who had previously interacted with a game companion. In this study, *The Wizard* was able to input either a dialogue act or a certain keyword shown in the screen to let the system decide automatically which behaviour to perform as a feedback to the player during the interaction.



Figure 5.4: Pupils interacting with Poppy at Firbeck Academy.

The first public event where the system was deployed was the MayFest 2015 on the Saturday 9 May 2015. In this event, there were 46 distinct participants (25 males, 34 children) with 50 interactions (44 with Poppy, and 6 with Spike). The total duration of interactions was 2 hours 32 minutes and 12 seconds ($MIN = 35s; MAX = 559s; AVG = 181.94s; SD = 90s$). All participants spontaneously played with the game companions without knowing what they should/can do with the virtual game companions except for 1 participant who had previously interacted with a game companion. From all participants, 32 participants finished their game with Poppy (winning rate 75 %). All participants who played with Spike did not want to continue playing with Spike because it was too hard to make Spike laugh and ultimately got bored and left the game companion immediately. Participants showed a negative emotion conveyed through their facial expression when interacting with Spike (see Figure 5.5). Moreover, some children players left the game companions in the middle of the interaction as soon as their parents or relatives or friends called them out to leave. The details of the second evaluation are described in the next section.

5.5.2 Results

Deploying virtual human in a public space is quite tricky to be implemented. The best position of both the system and participants is also potentially generated problems in the

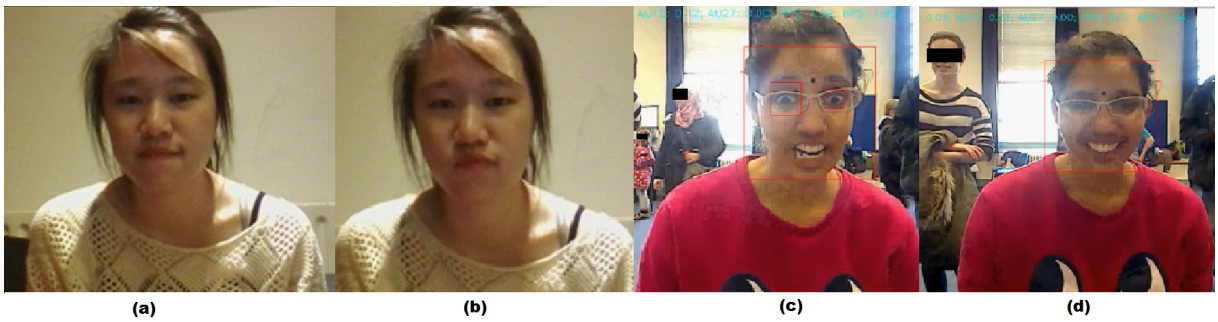


Figure 5.5: Player's Facial Expressions: Annoyed by Max in The Skyrim Game (a) and (b); Irritated by Spike (c) and Delighted by Poppy (d) in The Smile Game

interactions. The problem encountered in this study was that due to the location where the virtual humans were deployed, in an extremely noisy public location. This triggers two problems in the interactions, recognition in speech and facial expressions. It is almost impossible to reduce the background noise to improve the speech recognition as the background noise was extremely loud. Moreover, Participants sometimes had to move closer to the speaker as they could not hear the virtual human's voice. Hence the camera could only see part of the participant's faces. Changes in participant's positions (e.g. their location, either closer, farther or out of the frame) also sometimes causing a recognition problem. For example, in the MayFest study, some participants (11.54%) had their friends or family to accompany them when they were playing with the virtual human. Resulting in their faces overlapping each other in one frame. However, those participants were also having fun playing not only with the virtual human but also with their friends or family. Finally, almost 10% of the participants were playing with snacks in their hand (e.g. ice cream or candy), where sometimes causing an occlusion on their faces particularly around their mouth area.

Topic Analysis

Table 5.6 illustrates the conversation topics between players and the game companions during the interaction in the MayFest 2015 study annotated from the interaction videos (2560 x 1024 and 20 fps). The favourite topic to talk about during the interaction was about the game, with 352 occurrences in a total of average time of 1128.97 s (about 3.21 s/turn). The previous studies in the semi-public space indicated that the topic about news was not a good topic to

start the interaction. There were 83 % participants who turned the conversation down with saying "no good news". Thus, in this study, a context-specific topic (i.e. about The Mayfest event) (e.g. "Did you enjoy Mayfest?" or "What is your favourite activity in Mayfest?") was introduced. Participants responded positively about this topic with total AVG duration of 305.72 sec spent on this topic in 201 occurrences. This made the topic about MayFest become the second favourite conversation topics in this study. Moreover, some players were interested in the game companion's location when they noticed the picture of the mountains behind the game companions. A total of 99.59 s of the conversation time spent to discuss the game companion's location.

Table 5.6: Topics in The MayFest Study

Annotation	Occurrences	Total AVG Time(s)
About Agent	**38**	**57.33**
ask-wh-question-aboutagent	18	15.91
statement-opinion-aboutagent	20	41.42
About Game	**352**	**1128.97**
answer-yes-no-question-readytoplay	62	131.41
ask-wh-question-aboutgame	1	17.32
ask-wh-question-aboutgamerules	54	240.67
ask-wh-question-aboutgamestrategy	5	14.25
ask-yes-no-question-playgame	97	389.19
ask-yes-no-question-stopgame	6	15.38
ask-yes-question-aboutgamestrategy	17	28.74
statement-opinion-aboutgame	104	287.57
statement-opinion-aboutgamerules	3	4.44
About MayFest	**201**	**305.72**
statement-opinion-aboutmayfest	66	112.55
ask-wh-question-aboutmayfest	90	122.23
statement-opinion-aboutmayfest	5	7.22

Table 5.6: Topics in The MayFest Study

Annotation	Occurrences	Total AVG Time(s)
statement-opinion-aboutmayfest-positive	38	59.77
statement-opinion-aboutmayfest-negative	2	3.95
About Other Agent	**12**	**25.23**
ask-wh-question-aboutotheragent	12	25.23
About Place	**18**	**99.59**
answer-wh-question-aboutplace	2	4.10
ask-wh-question-aboutplace	3	11.93
statement-opinion-aboutplace	13	83.56
About Player	**41**	**69.79**
statement-nonopinion-aboutplayer	41	69.79
Clarification	**36**	**69.77**
ask-clarification	4	11.93
ask-to-repeat	30	53.60
apology	2	4.24
Greetings	**164**	**385.68**
greeting	164	385.68
Goodbye	**35**	**105.03**
goodbye	35	105.03
Error	**2**	**8.35**
statement-nonopinion-confusion	1	2.45
statement-nonopinion-wizarderror	1	5.90
Others	**4**	**6.77**
ask-yes-no-question-aboutturingtest	1	2.01
ask-yes-no-question-aboutspeeddating	1	4.76

Attack Analysis

In this study, facial expressions once again become the most popular choice for the players to attack the game companions. A total of 184 expressions were performed by the players in a total of average time of 139 s. Moreover, jokes come up in the second place as the choice of the strategy to make the game companions laugh with a total of 76 jokes performed in a total of average time of 92 s (see Table 5.9). Finally, As a result of the experimental setting of the MayFest, most of the participants were new to the system. Therefore, this setting did not allow us to evaluate the social relationship modules.

5.6 The Smile Game in Public Space: The Open Days

The University of Nottingham (UoN) Open Days 2015 was the second public event where The Smile Game deployed. The UoN OpenDays is an event held to provide the information to prospective students about the courses offered by the university. The target market for OpenDays are usually teenagers and their parents. Similar to the Mayfest 2015 study, any person of any age who showed an interest was invited to play with the characters by the attending researcher. All interactions were video and audio recorded. To adhere with the research ethics procedure, a consent from the parents or guardians was required if the participants were minors (aged 17 and below). Moreover, it was also not possible to run the system in a fully automatic mode due to the extremely noisy environment in the area where the system installed. Hence, the Semi-Automatic Wizard of Oz system similar to the Mayfest 2015 setting was applied in this study.

5.6.1 Experimental Setup

This study was conducted in an annual public event held by The University of Nottingham, the Open Days 2015. In this study, the system was running for two days, approximately for five hours per day. The study has also targeted any person of any age who were showing their

interest in the game. All the setting and evaluation methods were identical with the MayFest Study. This study also aimed to evaluate the virtual humans' system as well as the second version of computational models in the wild (i.e. public area). Due to the nature of the interaction (i.e. with mostly new and non-recurring person), it was not possible to evaluate the model of social relationship. Moreover, similar to the Mayfest 2015 study, due to the very noise background, the Automatic Speech Recognition and Verbal Interpreter Modules were failed to run automatically, thus semi-automatic with WoZ system was activated.

In total, there were seven distinct players (all males, three teenagers) with eight interactions, played with both characters (6 with Poppy and 2 with Spike). Both of the characters were interacted with all the players for a total of 20 minutes and 50 seconds ($MIN = 59s; MAX = 366s; AVG = 156.25s; SD = 92s$). There was a player who was invited to interact with our system, while the others were spontaneously interacted with the system without knowing anything about the system, game and the game companions. All participants were interacted with the system from the beginning until the end of the phase, and half of them won the game against the Poppy and no one won against Spike (winning rate 50 %). All players who played with Spike perceived Spike as someone who is too serious to play the game and claimed that no one would be able to beat the character. Identical with the studies in The *Semi-Public* Space settings, all the studies were were video and audio recorded from the perspective of both the player and the NPC to be further annotated (See Figure 5.2).

5.6.2 Results

The problem arose in the Open Days 2015 study was in the video input due to the system was installed in a location with excessive natural light during the study hence disrupting the recognition systems from the video input (i.e. camera). Moreover, similar to the Mayfest 2015 study, noise was also a critical problem in this study. This triggers two problems in the interactions, recognition in speech and facial expressions. It is almost impossible to reduce the background noise to improve the speech recognition as the background noise was extremely loud. Moreover, Participants sometimes had to move closer to the speaker as

they could not hear the virtual human's voice. Hence the camera could only see part of the participant's faces. In contrast with the Mayfest 2015 study, in this study, there were no participants interacting with the system with snacks in their hand and mouth. Hence, there was no any serious occlusion on their faces, particularly around their mouth area.

Topic Analysis

Similar with the MayFest 2015 study, in the Open Day study topics about the game and the event were the most favourite topics between players and the game companions during the interaction. Table 5.7 demonstrates the annotation of the conversation topics from the videos (2560 x 1024 and 10 fps). There were 94 and 40 occurrences in a total of average time of 293.42 s and 64.52 s spent in the game and Opendays topics respectively.

Table 5.7: Topics in The OpenDays Study

Annotation	Occurrences	Total AVG Time(s)
About Agent	**21**	**128.32**
ask-wh-question-aboutagent	19	125.77
statement-opinion-aboutagent	2	2.55
About Game	**94**	**293.42**
answer-yes-no-question-playagain	7	16.97
answer-yes-no-question-playgame	14	20.44
ask-wh-question-aboutgame	20	71.25
ask-wh-question-aboutgamerules	15	57.01
ask-yes-no-question-playgame	8	60.58
statement-opinion-aboutgamerules	6	11.58
statement-opinion-aboutgame	24	55.59
About OpenDays	**40**	**64.52**
ask-wh-question-aboutopendays	16	26.32
statement-opinion-aboutopendays	24	38.20

Table 5.7: Topics in The OpenDays Study

Annotation	Occurrences	Total AVG Time(s)
About News	**12**	**25.94**
ask-wh-question-aboutnews	12	25.94
Greetings	**34**	**79.50**
greeting	34	79.50
Goodbye	**6**	**15.93**
goodbye	6	15.93

Attack Analysis

In this study, there were only three types of movements to attack the game companions performed by the players (see Table 5.9). Facial expressions still in the first place as a choice of the strategy to win the game, with 40 occurrences performed in a total of an average of 31 s. In the second place, a combination of facial expressions and hand movements was performed to attack the game companions. In contrast with all the studies, there were no jokes told in this study as a choice of the strategy to attack the game companions.

5.7 Semi-Public vs. Public Space

The interaction video inferred that most of the players who interacted with both characters, perceived Poppy as a nice and friendly person, while Spike was perceived to be a rigid, unpleasant character who was hard to defeat in the game, in both semi-public and public area. Figure 5.5 demonstrates one of the player's expression when playing with Spike (c) and Poppy (d) in The MayFest 2015 study and Figure 5.4 illustrates pupils of the Firbeck Academy was interacting with Poppy (image was taken from The University's news website[2]).

[2]http://exchange.nottingham.ac.uk/blog/puzzles-robots-and-raspberry-pi-computer-science-activity-day-at-firbeck-academy/

Most of the players irritated by Spike when playing with him and some of them did not wish to continue to interact with him. A player in the MayFest 2015 study stated "I don't want to play with him any more, He is so mean." and another player described Poppy and Spike with a statement "[...] well, she definitely looks prettier than that gentleman. He looks scary. (laugh)" to her friend in the MayFest 2015 study. Another player also said "I don't want to play with you (act crying)" to Spike in the middle of the game. The researcher is also aware that the characters voice and face features also provide an extra effect to the personality perception by the players. Together with the framework and models proposed and good game design, this provided an emotionally immersed and enjoyable game (Lazzaro 2004).

Topic Analysis

Table 5.8 shows a summary of top 5 topics that players like to talk about, with the virtual game companions in all studies. In general, the topic about the game itself was the most popular topic for conversation with the virtual game companions with total occurrences of 589 times and total average of the duration of 1716 s (2.91 s/turn). In the fully automatic study, topics related to clarification utterances and errors were spotted in the top 5 topics, although there were also some error and clarification related topics in the other studies. This was the results of the failure of the Automatic Speech Recognition module in the study, and the system failed to interpret what the players were saying during the interaction correctly. In the semi-public space, most of the players turned down the conversation when the game companions started a topic about news. The players either said there was no news that players can tell or they have not read or heard any news that day. Hence, even though the game companions already initiated the conversation by giving the news of the day, the players seemed did not interested in continuing the topic. Therefore, in the public space studies, the utterances about the news topic were replaced by a topic related to the events, such as what kind of activity that the players like in the event. This was positively received by the players. The players seemed to enthusiastic to talk about their activity in the event. The conversation flowed smoothly in this topic compare to the topic about news, with a

total of 241 times occurred in both MayFest and OpenDay 2015 study with a total of 371 s average time (about 1.54 s/turn).

Table 5.8: Top 5 Topics in The Studies

Study	Topics	Occurrences	Total AVG Time(s)
I	About Game	92	178
	About Agent	25	65
	About News	21	54
	Clarification	19	19
	About Life	12	22
II	About Game	51	116
	Clarification	45	40
	About News	43	70
	About Presence	12	21
	Error	9	32
III	About Game	352	1129
	About Mayfest	201	306
	About Player	41	70
	About Agent	38	57
	Clarification	36	70
IV	About Game	94	293
	About Opendays	40	65
	About Agent	21	128
	About News	12	26

Attack Analysis

Table 5.9 illustrates top 5 movements performed in all studies. In most cases, attacks with facial expressions were the most popular movement across all the studies. In total, the players performed 284 attacks with only their facial expression in a total of average time of 206 s (around 0.73 s/attack). This is a good sign, as the system was only capable to automatically detect an attack performed by the player' facial expressions through the activation of their facial Action Units. An attack with jokes also quite popular among the other movements. There were a total of 111 jokes expressed in the game in a total of average time of 248 s (about 2.23 s/attack). In the third study (i.e. The MayFest 2015), the "knock-knock" jokes were favourite among the other jokes. Moreover, players also combined their facial expressions with others movements such as body or head movements as an attack strategy. Uniquely, no players in the fourth study (i.e. The OpenDay 2015) were performing jokes as

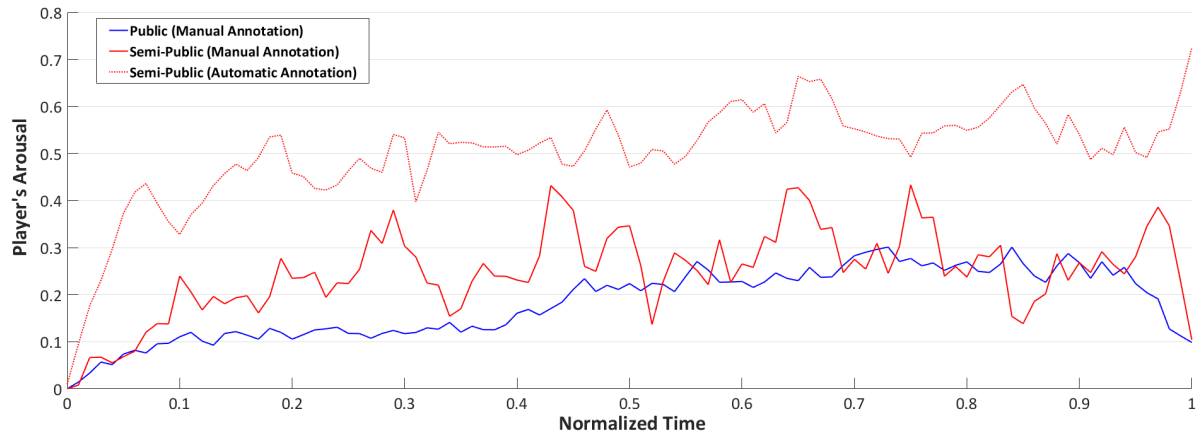


Figure 5.6: Players' Arousal During the Smile Game in *Semi-Public* (red) and *Public Space* (blue), where dotted line represents Automatic Annotation

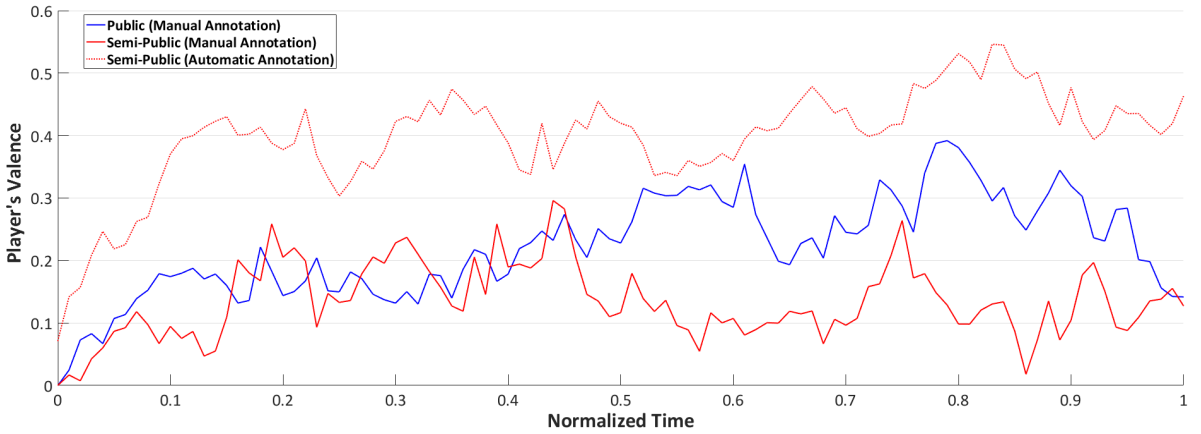


Figure 5.7: Players' Valence During the Smile Game in *Semi-Public* (red) and *Public Space* (blue), where dotted line represents Automatic Annotation

their attack strategy to win the game. Most of them were only relied on their facial muscles to defeat the game companions.

Arousal and Valence Analysis

Furthermore, automatic arousal and valence annotation were performed from the interaction audio using the SSI Framework (Wagner et al. 2013). The automatic annotation was only performed to the interactions audio from the *semi-public* study, as the one from the *public* space setting was too noise too automatically annotated. In addition, we also manually annotated players' arousal and valence using continuous annotation tools developed by Imperial College London. There was a moderate positive correlation between the data from auto-

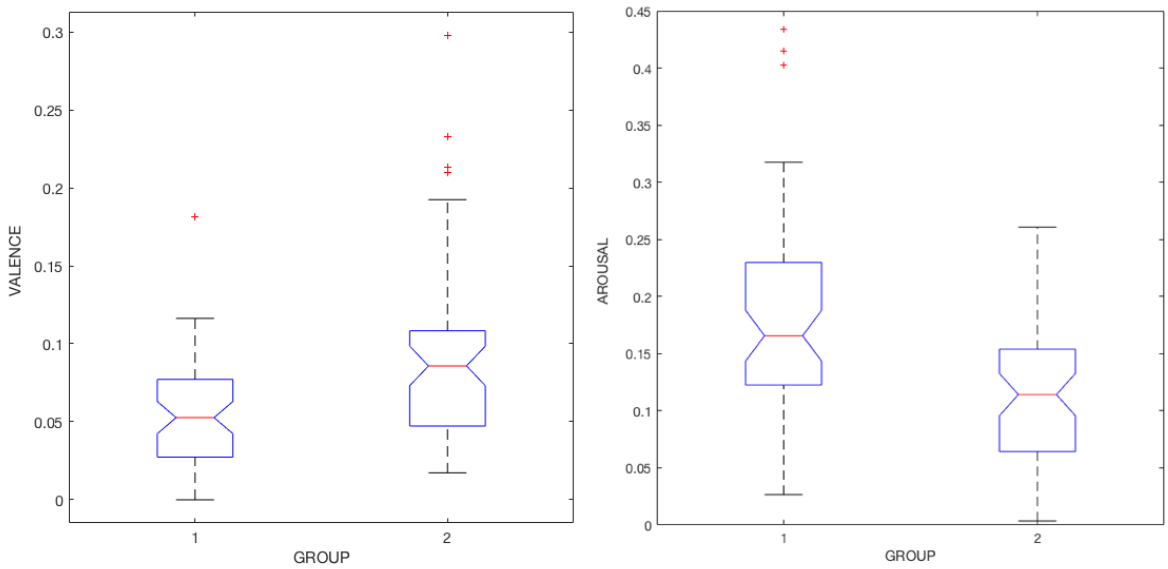


Figure 5.8: Player's Valence (Left) and arousal (Right) during the Smile Game in *Semi-Public* (Group 1) and *Public* (Group 2) Space) form the Manual Annotation

matic and manual annotation, $r = 0.66, n = 101, p < 0.01$ and $r = 0.51, n = 101, p < 0.01$ for arousal and valence respectively. In addition, the Kolmogorov-Smirnov test was applied to all the data and indicated that they were more likely to come from a non-normally distributed population ($p < 0.001$). Hence Wilcoxon Signed-Rank Tests were applied as this is a good alternative to the t-test Paired Two Sample for Means when the population cannot be assumed to be normally distributed.

Figure 5.6 and Figure 5.7 illustrates the players' arousal and valence averaged over all players, during playing with both characters. As the interaction durations are distinct from each other, the time data were normalised, where 0 indicates the beginning and 1 refers to the end of the interaction. The peak of players' arousal and valence was seen in the almost end of the interaction before it went down approaching zero at the very end of the interaction. The interaction reached its climax when the players showed their best effort to make the ECAs smile as the music indicated that the game was approaching its end.

Moreover, a Wilcoxon Signed-Ranks Test indicated that players' arousal from the manual annotation was statically significantly higher in the *semi-public* space compare to the one in the *public* area, $Z = 7.3, p < 0.001$. This indicates the players were more engaged in a *semi-public* space. In contrast, players' valence was statistically significantly higher in the

public space, $Z = 5.73, p < 0.001$. This indicates that, in this game, players showed more positive expressions in a *public* area, possibly exaggerating their signals of enjoyment. Figure 5.8 shows ANOVA analysis of player's Valence and Arousal (from the manual annotation) during the game in *Semi-Public* and *Public* Space. The ANOVA analysis also indicates that the player's arousal in the *semi-public* space was statically significantly higher than the *public* area ($p < 0.001$), while player's valence was statistically significantly higher in the *public* space ($p < 0.001$).

This could be explained by the fact that in the public area the players were in a social environment, often with their families or friends. Research has shown that players are having the most fun when playing with friends and they are more emotionally expressive when playing with friends (Shahid et al. 2011, 2012). In addition, the players' valence and arousal increased at the beginning of the game and remained relatively constant in the middle throughout to the end of the game in both spaces, except for the valence in a *semi-public* space. In a *semi-public* space, players' valence was slightly decreasing in the middle of the game, indicating the decline of the players' enjoyment.

Table 5.9: Top 5 Movements in The Studies

Study	**Movement Type**	**Occurrences**	**Total AVG Time(s)**
I	Facial Expressions	34	30
	Jokes	26	29
	Body Movements + Facial Expressions	7	17
	Body Movements	5	11
	Head Movements	3	5
II	Facial Expressions	26	6
	Facial Expressions + Body Movements	1	4
	Jokes	9	3
	Head Movements	4	3
	Facial Expressions + Head Movements	2	2
III	Facial Expressions	184	139
	Facial Expressions + Hand Movements	63	105
	Jokes	76	92
	Head Movements	21	33
	Noise	16	19
IV	Facial Expressions	40	31
	Facial Expressions + Hand Movements	5	18
	Head Movements	1	2

5.8 Conclusion

The studies demonstrate the implementation of a fully integrated and modular framework for social-affect-aware in a game in the wild, which focuses on the social interaction between the player and virtual human as well as the models for personality, emotions and social interaction (**RQ1**). Generally, people enjoyed to interact with the virtual human in all studies but the one with the Fully Automatic System in the Semi-Public Space. Overall, the interaction reached its climax when the participants performed their best effort to make the virtual human smile near the end of the interaction. From the valence and arousal annotation, the players were more engaged in a *semi-public* space. Most likely that the players were more confident to shows their silliest facial expressions in *semi-public* space. However, the player's showed more positive expressions in a *public* space, possibly exaggerating their signals of enjoyment, where some participants played with their friends or families as well during the game.

Overall, the players perceived Poppy as a nice and friendly person, while Spike is a rigid and unpleasant character who was hard to defeat in the game. This shows that personality also could be implemented as the game balancing and levelling system. From all studies, players were interested in talking about specific things related to the game and event, for example: in The MayFest study, topics regarding the MayFest event occurred 201 times in total. Other than that, the topic about the Player and the Virtual human also interest the players. Moreover, facial expressions attack was the most popular action performed by the players during the game. Hopefully, the results of favourite topics and attack could be used to reduce the uncertainty of the possibilities of the interaction between human and virtual human in the wild (**RQ5**).

There are several things required to be taken into consideration when developing a system with similar settings to ours. First, when having a virtual human that can interact in multiple phases of interactions (e.g. in this thesis: introduction, chit-chat, game and closing), a definite turn taking should be apparent. Moreover, the participants should be able to differentiate one phase from the other phase as well as aware that the interaction phase has

changed. In this research, we implement music to differentiate the interaction phase and use several turn-taking cues to distinguish the turn taking. The position of the system also should be taken care, as in public space, there are several noises to the sensors, such as lightning to a camera and background noise to a microphone.

From the technical system evaluation perspective, the virtual human is reasonably good in perceiving player's facial expressions especially for AU 12 (Lip corner puller), a bit slow but reasonably accurate for AU 27 (Mouth stretch), but sometimes give false positive for AU 1 (Inner brow Raiser). Based on the debug screen (See Fig 5.2), the companion's urge to laugh (emotion perception) is reasonably good in reacting to the player's action from facial expressions. However, the limitation of this system is that the companions were only capable of perceiving a small number of cues from facial expressions and only capable of perceiving verbal signals from the audio. Moreover, to answer **RQ4**, this chapter also presents on how to collect and craft the interaction rules for virtual human in a specific game scenario.

The virtual human's average recognition rate was 89.19% in all studies where the players interacted with the system more than once. As a result of the experimental setting of the MayFest and OpenDay 2015 and a relatively low turnout of the other studies, most of the players were new to the system. Therefore, this setting did not allow us to evaluate the social relationship model. Hence, The Skyrim Game Study was designed to evaluate all the models and framework including the social relationship model in a more controlled setting. Moreover, in the fully automatic system, the failure of the virtual game companion's perception of the verbal communication is a substantial obstacle to overcome. Seven out of 12 players were not able to play with the virtual human because the companion failed to understand the player's utterances and repeatedly asked the participants to rephrase their word or talked randomly.

Chapter 6

Augmenting Existing Game Characters using ERiSA

6.1 Introduction

This chapter presents the implementation of the ERiSA framework and models in an existing commercial game, Skyrim. In contrast to The Smile Game, where the human players and the game companions are not inhabiting the same world (the player inhabit the real world while the game companions inhabit a virtual world), in The Skyrim Game, both human players and game companions are inhabiting the same, virtual, world. The ability of players and NPC's actions give us the opportunity to focus on the evaluation of the models and parts of the system that were found to be promising in Chapter 5.

A study with Action Role-Playing Game (RPG) was designed to evaluate player interactions with game companions, who were imbued with social and emotional skill by the ERiSA Framework together with the computational models. Players had to complete a quest in the Skyrim game, in which players had to use social and emotional skills to obtain a sword. The results clearly show that game companions who are capable of perceiving and exhibit emotions, are perceived to have personality and can forge relationships with the players, enhancing the player experience during the game. Players reported that they were more

emotionally involved and attached to the characters and the game when the Framework was activated in the game.

6.2 The Skyrim Game

We applied the second version of the computational models to The ERiSA Framework described in Section 3.3 and Skyrim Creation Kit[1]. Figure 6.1 illustrates the integration of the components and the communication between components. Parts of the ERiSA Framework are used to perceive and interpret the player's emotions, conveyed by their facial expression (Valstar 2015, Almaev et al. 2013) and recorded by a camera. The ERISA's Interpreter Components interpret the information captured from the player's facial expressions by using a web camera. The interpreted emotions sent to the Dialogue Management Module in the Behaviour Components. We integrated the ERISA's Behaviour Components to the Skyrim Creation Kit Dialogue Manager. The combined Behaviour Components then propose the game companion's actions and a set of possible dialogues for the player to choose depending on the player's and game companion's internal states. Then the output is sent to the Skyrim Game Engine. A player's facial expressions also influence the player's options for dialogue when interacting with the game companion. For example, the option for the player saying happy is only available if the facial expression components interpret the player's emotion conveyed by their facial expression as happy.

In addition to the mechanics, stories, aesthetics and technologies provided by Skryim, we designed two characters and a quest to the game. Moreover, the ERiSA computational models modify the mechanics for building a relationship between player and game companion in the game. However, we used Skyrim's original mechanics for building relationships (see [2] and [3]) as a baseline to compare with. We argue that with our proposed computational models implemented to the Skyrim game will increase the player experience (e.g. emotions when playing games), immersion and engagement to the character, game and story.

[1] http://www.creationkit.com/
[2] http://www.creationkit.com/index.php?title=Relationship
[3] http://www.creationkit.com/index.php?title=Faction

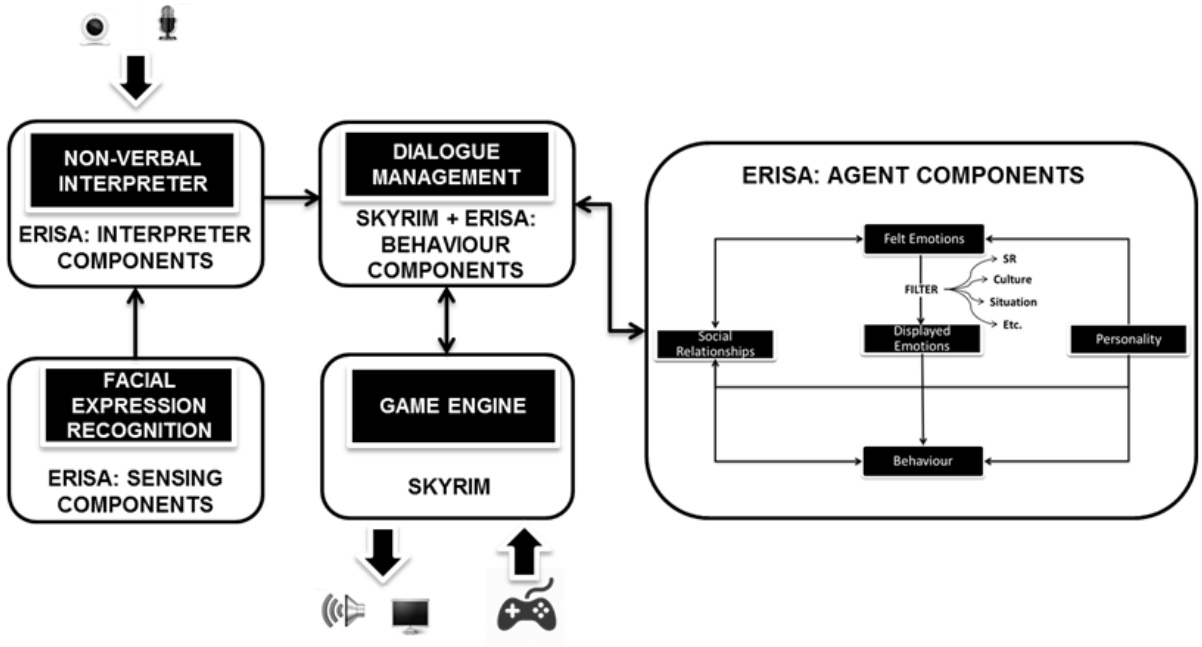


Figure 6.1: Framework Implementation

6.2.1 Experimental & Character Design

We designed two game companions with opposite personality to interact with (see Figure 6.2). We called them, Stella and Max Erisa. Two experimental scenarios were designed for each character for the models' evaluation. The first is a scenario with baseline characters that were implemented using the original mechanics of the Skyrim Game (see [4]) and the second scenario with enhanced mechanics using the ERiSA models and framework. The quest for both baseline and augmented character scenarios and the methods available to a player to finish the quest are identical, only the characters' behaviour models are different. Table 6.1 illustrates the differences between baseline and enhanced characters.

Baseline Characters In this scenario, the characters are designed with the Skyrim Game mechanics where the NPC's emotions and social relationships are triggered by the story or click-based actions. Moreover, as the ERiSA framework was not implemented for these characters, the dialogues between player and NPC are only regulated by the Skyrim Dialogue Manager, which does not include the emotions conveyed by the player's facial expressions in the decision making. As it is implementing the conventional click based system from the

[4] http://www.creationkit.com/index.php?title=Category:Game_Systems



Figure 6.2: Character Design

Skyrim Game original mechanics, technically, the players can finish the quest in four clicks if they would choose the right sequence of dialogue options.

Enhanced Characters In this scenario, we designed Stella to be a high Extraversion person ($P_{E,N} = \{0.5, 0\}$, $\theta_{v-} = -0.3$, $\theta_v = 0.2$), while Max has high Neuroticism personality trait ($P_{E,N} = \{0, 0.5\}$, $\theta_{v-} = -0.2$, $\theta_v = 0.3$). To associate the characters with their personality traits, we set Max voice characteristic to an agitated and annoying voice provided by the Skyrim Creation Kit library. On the other hand, Stella voice is pleasantly happy, and enthusiastic. In addition, the behaviours of the game companions (e.g. choices of the words) are characterized by the pattern presented in Table 3.1 to match the characthers' personality. The other independent variables of the characters are emotion decay rate $r = -0.5$, $\mathcal{Y} = 200$, $\mathcal{W}_p = 0.2$, SR decay rate $r = -0.25$ and $\mathcal{S}_{i,j} = 0.02$ for both characters. The NPCs' emotions and their ability to forge relationships with players are governed by the ERiSA models specified in Chapter 3.

Table 6.1: Baseline VS. Enhanced Characters

Character	Components	Baseline	Enhanced
Max	Appearance	Blacksmith with armor	Blacksmith with armor
	Voice	Agitated, annoying	Agitated, annoying
	Quest	Legendary sword	Legendary sword
	SR Mechanism	Skyrim	Skyrim + ERISA
	Personality System	Skyrim	Skyrim + ERISA
	Emotions Regulation	No	ERISA
	Affect Aware System	No	ERISA
Stella	Appearance	Alchemist with robe	Alchemist with robe
	Voice	Happy, enthusiastic	Happy, enthusiastic
	Quest	Legendary sword	Legendary sword
	SR Mechanism	Skyrim	Skyrim + ERISA
	Personality System	Skyrim	Skyrim + ERISA
	Emotions Regulation	No	ERISA
	Affect Aware System	No	ERISA



Figure 6.3: Interaction with Max (left) and Stella (right) in the Beginning of the Quest

6.2.2 Story: The Erisa Family

In the game, we designed a simple quest named "The Erisa Family". In the quest, the player can build social relationships with the game companions, get information about The Legendary Swords possessed by their family and ultimately retrieve one of the swords. A player can dynamically interact with the game companions, which depends on their social relationship level. The quest is finished when the player gets one of the swords. The Sword of Friendship will be given to the player when the player can gain the game companions' trust by building a positive relationship, while The Sword of Hatred will be passed to the player when their social relationship is going towards a negative direction.

Stella and Max are brother and sister. Building a relationship with Stella is easier than with Max as Stella is an easy-going person. Stella is the oldest children of the family with a brother and sister. She works as an alchemist, which is why she likes alchemical ingredients such as potions and natural herbal ingredients. She loves her family; However, she is against the idea that women must stay at home and can't have a job. She argues that women should be able to do what they want to do. Max, on the other hand, hates everyone including his family. He particularly hates his father because in his opinion, he should not have retired from his prestigious job as a "Jarl" so soon and he should not now be doing nothing at home. "Jarl" is the highest rank in the Holds of Skyrim. There are nine Holds which regulate several cities around them. "Jarl" is also part of the Elder Council, they have the power to legislate for the entire Empire. Max currently works as a blacksmith in the town, which is why he likes metals such as Adamantium and Palladium. All of this information is slowly disclosed to the players by the game companions depending on their relationships and can be exploited to forge relationships with the game companions.

There are four significant strategies designed to forge relationships, either positive or negative, with the game companions: talking, giving favourite items vs giving undesirable items, praising vs criticising and exploring Skyrim world together vs attacking the game companions. Stella is more sociable, compared to Max. Hence, an extra effort is required to get Max's trust. Over time, the topics discussed with the game companions change depending on the level of social relationship between the player and the game companions. Figure 6.4 illustrates the mapping between possible interactions corresponding to the level of social relationship based on the ERiSA model (see Subsection 3.3.3), and Figure 6.4 shows the player's interaction with Max and Stella in the beginning of the quest.

There are four levels of social relationships for positive relationships and four levels for negative relationships. On the initial interaction, the player can only discuss general topics about the swords, city, Skyrim, giving items, asking for and giving news. Assuming the social relationship is going towards a positive direction, in the next stage, the game companions start to express their personal attitudes towards the topics. A moderate topic is available to discuss in this stage in addition to the previous topics. Starting from this stage, the player

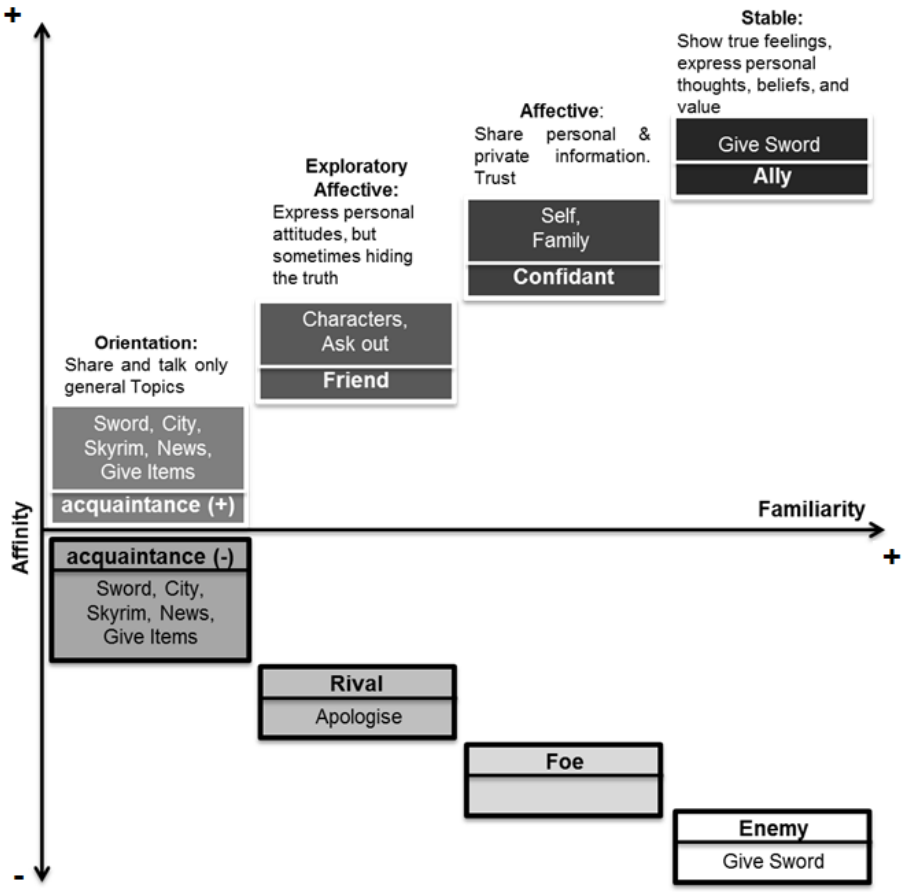


Figure 6.4: Social Interactions and Topics

can ask the game companion to go out and explore Skyrim together.

Going towards a high-affinity relationship, the game companions start to share their personal and private information about their self and their family. Moreover, the game companions strongly express their personal attitudes towards the topics at this stage. Finally, the game companions begin to show their true feelings, express personal thoughts, beliefs, and values. At this stage, the player is given The Sword of Friendships, and it marks the end of the quest. The player still can interact with the game companions, but the social relationship between them will not go any further. On the other hand, assuming the social relationship is going towards negative direction, the breadth and depth of topics do not change, but there is room to apologise to the game companions. Eventually, a Sword of Hatred is passed to the player marks the end of the journey. If this route is taken, the player will not be able to interact with the game companion any more, assuming the game companion holds a deep hatred to the player.

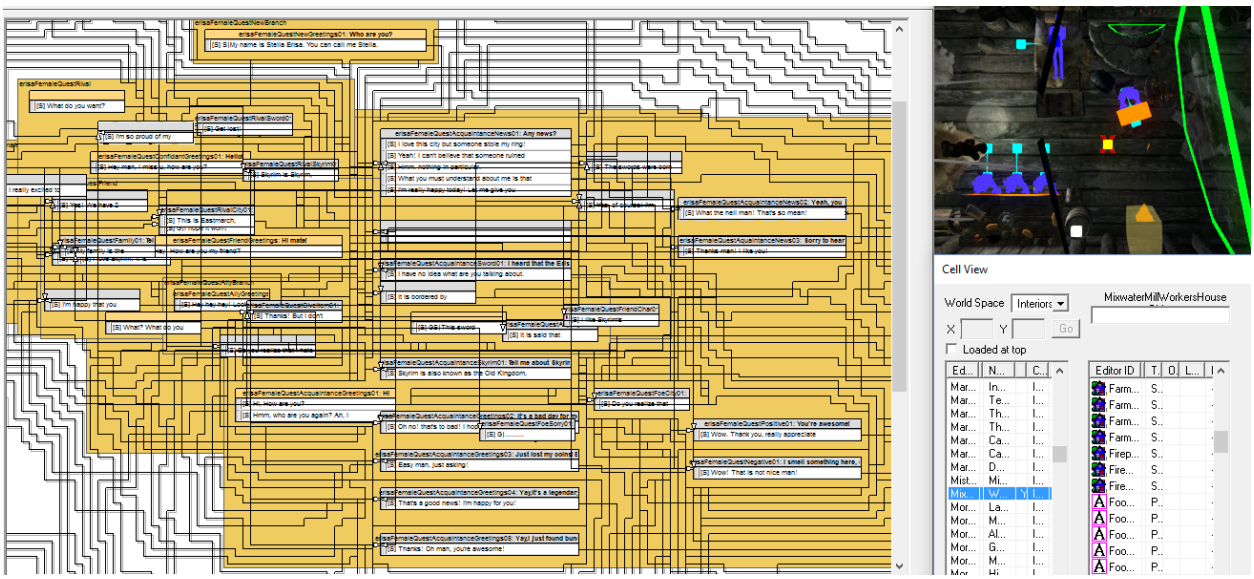


Figure 6.5: Skyrim Creation Kit

6.2.3 Dialogue Management

The dialogue system in Skyrim implements a dialogue tree model provided by Skyrim Creation Kit. Figure 6.5 illustrates the environment provided by the Skyrim Creation Tool kit to design the game dialogues. Ten topics were designed to the game dialogues. As described in the previous section, in the enhanced character scenario, the availability of particular topics depends on the social relationships between them, as well as the player's and game companion's emotions. To make it easy to maintain, we designed the dialogue with a card-based method. Each card contains information about the topic category, utterances, conditions and the effects. Figure 6.6 shows the example of the dialogue design. The conditions indicate the requirement criteria for the dialogue card to appear as a dialogue option for the player. The utterances show the reply that the player will get if they choose the utterance. Finally, each card has its own effect after it has been chosen. For example: a Sword01 card has a pre-condition that the level of their social relationship is acquaintance (either positive or negative), *Affinity* ($\mathcal{A}_{i,j,t}$) and *Familiarity* ($\mathcal{F}_{i,j,t}$) of the NPC i towards player j at time t must be in the low category. If those conditions are met, the option of a player is saying "I heard that the Erisa Family has the Legendary Swords" is available. Furthermore, it will give an increase to the NPC's *Familiarity* ($\mathcal{F}_{i,j,t}$), if the player chooses that utterance.

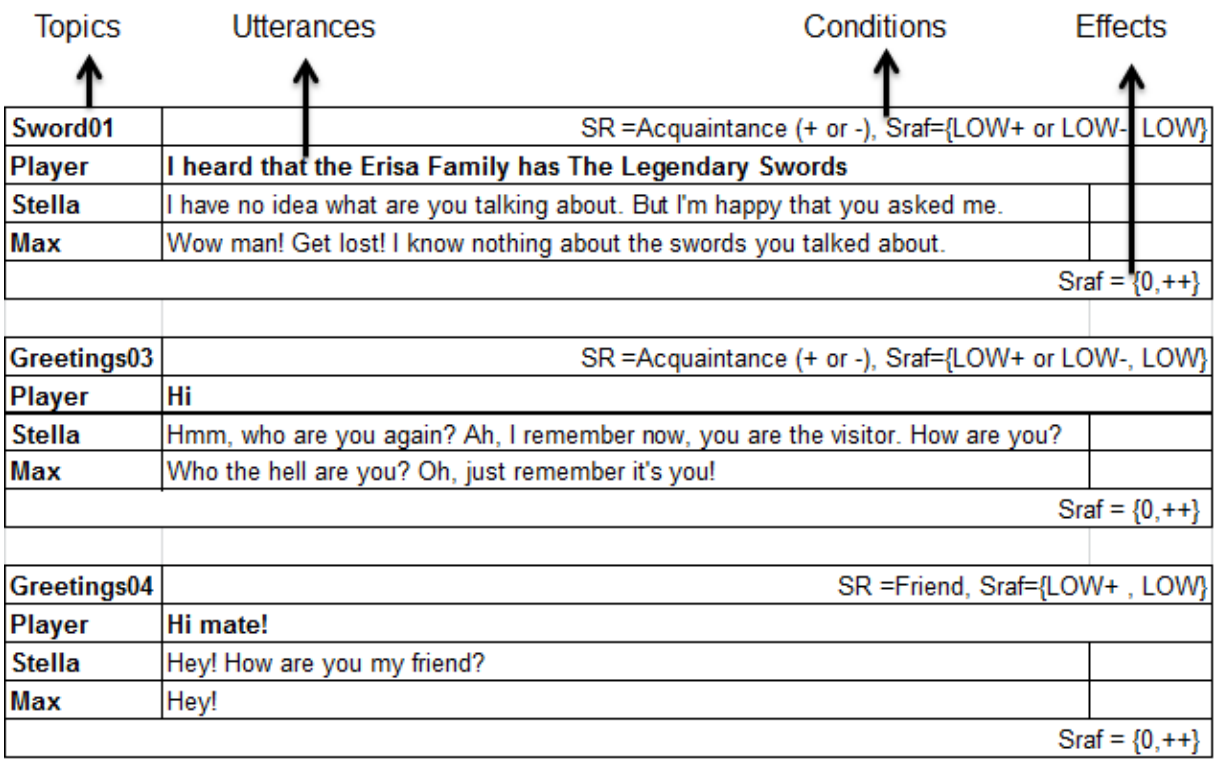


Figure 6.6: Skyrim Dialogue Management

6.3 The Experiment

6.3.1 Experimental Setup

To comprehensively evaluate the ERiSA models, a controlled study was conducted. The models and part of the framework were implemented in an existing commercial game, Skyrim. This particular game was chosen as the Skyrim world provides free-form interaction and exploration to the world. This was a perfect feature to test interactions with characters with the models implemented. Also, the Skyrim developers provide a tool-kit for other developers to alter the game objects. Hence the ERiSA models and framework were easily implemented into the game using the tool-kit. In this study, the players and NPC's actions were able to be limited by design. Consequently, the researcher has a more controlled environment for the study. Because ASR was found to be a bottleneck in Chapter 5, instead of using the Automatic Speech Recognition module to recognise the player's utterance, the researcher

provided a list of utterances that the player can choose from based on the player's emotions, NPC's emotions and their social relationships.

In this study, fifteen participants were recruited (40% male, 73.33 % Asian and others were Caucasian, AVG age = 24.6) from the University of Nottingham to play the game. All of them are familiar with RPG or Action Game Genre, and one-third of them have played Skyrim before. There are a total of 60 interactions with both of the NPCs, with a total play time of 18.16 hours ($AVG = 18.16mins, MAX = 51.82mins, MIN = 1.85mins$). Each participant interacted with both NPCs, the baseline characters, and with the enhanced characters implemented as the comparison. To avoid order bias, the sessions were randomised for all the participants. All interactions were video and audio recorded using a web camera, the player and NPCs' internal states were logged, and the player's choices were recorded. The participants were asked to fill out a questionnaire every time they finished a session of the game. Moreover, some participants were invited to do a short interview (less than 5 minutes) to discuss their experience with the game, after the participants had finished all the quests.

The five-point Likert scale questionnaire consists of four parts of evaluation towards the game and NPCs:

Percieved Characters Personality: The first part evaluates the NPC's personality perceived by the user based on the Big Five Personality Inventory (John et al. 2008). The participants were asked if they saw the NPC as people **P1**: with positive thinking, **P2**: with negative thinking, **P3**: who generates a lot of enthusiasm, **P4**: who tend to be quiet, **P5**: who is outgoing and sociable, **P6**: who is depressed and blue, **P7**: who is relaxed and handle stress well, **P8**: who can be tense, **P9**: who worries a lot, **P10**: who is emotionally stable and not easily upset, **P11**: who can be moody, and **P12**: who is sometimes rude to others.

Percieved Social Relationship: Second part is to rate the social relationships between the participants and the NPCs, where **R1**: [NPC name] is someone that you can count on to listen when you are very angry at someone else, **R2**: I really like [NPC name], **R3**: I'm familiar with [NPC name], and **R4**: I am close with [NPC name].

Player's Self-Reported Emotions: In the third part of the questionnaire, the participants were asked about their feelings when they interacted with the NPC during the quest, with **E1**: Happy, **E2**: Angry, **E3**: Sad, **E4**: Afraid, **E5**: Disgust, and **E6**: Surprise.

Game Experience: The last part of the questionnaire evaluated user experience during the game based on the Game Engagement Questionnaire (Brockmyer et al. 2009) consists of **G1**: Playing seems automatic, **G2**: I lose track of where I am, **G3**: I play without thinking about how to play, **G4**: I feel like I just can't stop playing, **G5**: I really get into the game, **G6**: If someone talks to me, I don't hear them, and the Immersive Experience Questionnaire (Jennett et al. 2008) where **G7**: I felt that I really empathised/felt for with the game, **G8**: I did not feel any emotional attachment to the game, **G9**: I was interested in seeing how the game's events would progress, finally **G10**: I sometimes found myself to become so involved with the game that I wanted to speak to the game directly.

6.3.2 Results

Player's Emotions

Shapiro-Wilk test for normality (Shapiro & Wilk 1965) was applied to all the data distribution with more than two third of the data (67.5%) are likely to come from a non-normally distributed population ($p < 0.05$). Hence, we applied a Wilcoxon signed-rank test as an alternative to the t-test paired two sample for means, when the population cannot be assumed to be normally distributed. Figure 6.7 illustrates the player's self-reported emotions during the interaction with both characters in the game. In the interaction with Stella, the players reported a significant increase in their happiness ($p < 0.01$) in the game with models implemented. On the other hand, in the interaction with Max the players reported a significant changes in almost all their emotions with a decrease in happiness ($p < 0.01$) and an increase in their anger ($p < 0.01$), sadness ($p < 0.01$), afraid ($p < 0.05$) and disgust ($p < 0.01$) emotions with the character that models implemented in the game.

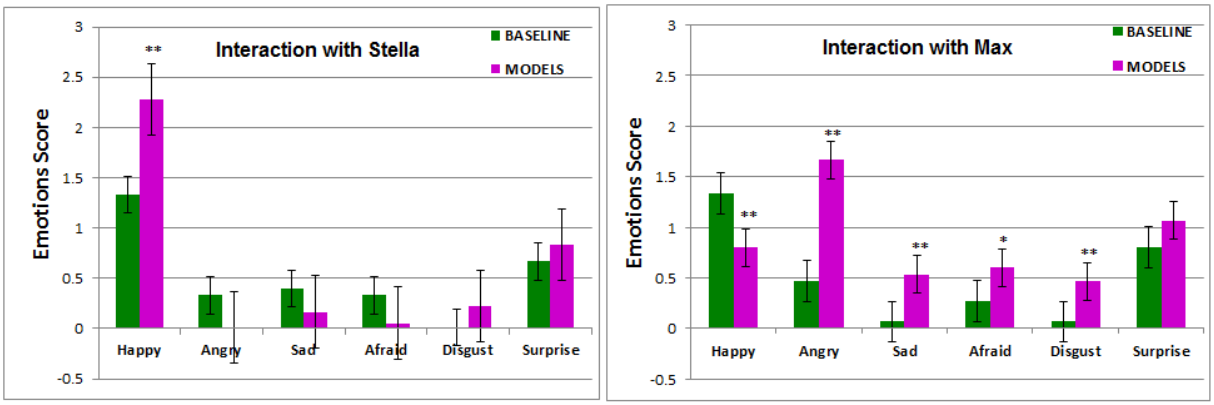


Figure 6.7: Player's self-reported emotions during interaction with Stella (left) and Max (right), with ** mark represents significance at the 0.01 level and the other mark * indicates significance at the 0.05 level.

Personality

Figure 6.8 describes Max and Stella's personality as perceived by players for both models. Without the models implemented, the NPCs' appear to be perceived as devoid of any personality by the players. With the models implemented, the characteristics which constitute Stella personality, are evidently perceived by players as a person who has positive thinking, generates a lot of enthusiasm, is outspoken, outgoing and sociable, is relaxed and handles stress well, is emotionally stable and not easily upset, and is polite to others. Max, on the other hand, is a person who is extremely negative thinking, does not generate a lot of enthusiasm, tends to be slightly quiet, is reserved, sometimes depressed, blue, and can't handle stress well, can be tense and worries a lot, is easily upset, can be moody, and is sometimes rude to others. In addition, most players identified Max as rude, abominable and playing hard to get, while Stella is a talkative, nice and approachable person. According to responses to the Game Engagement and Immersive Experience Questionnaire (see Fig 6.10), the players empathise with Stella but not Max in the game with models implemented, while there was no significant difference when models were not implemented. Fig 5.5 (a) and (b) shows one of the player's facial expressions when playing with the models empowered Max.

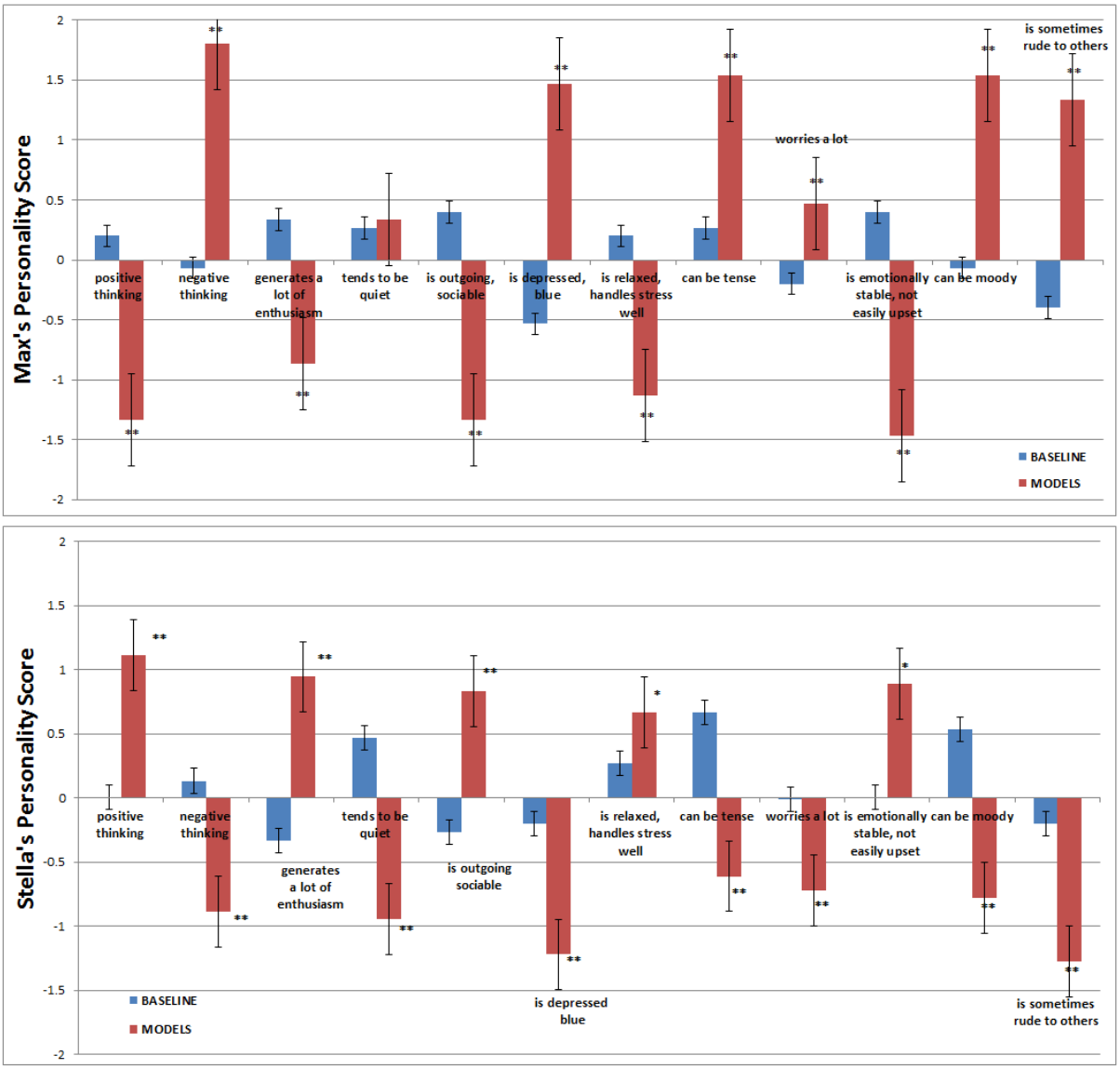


Figure 6.8: Max's (top) and Stella's (bottom) Personality as Perceived by Players with * mark represent significance at the 0.05 level and others marked with ** indicates significance at the 0.01 level

Social Relationship

Furthermore, players reported significant changes in their social relationship to both NPCs (see Figure 6.9). With the models implemented, Stella is more likeable, and players perceived her as someone that they can count on to listen when they are very angry with someone else. Players seem happy when interacted with her. While, Max is more unpleasant to interact with, and players feel he is someone that they cannot count on to listen when they are very angry with someone else as they already feel emotionally upset to Max. Players are also

more familiar and close with both NPCs when the models are implemented.

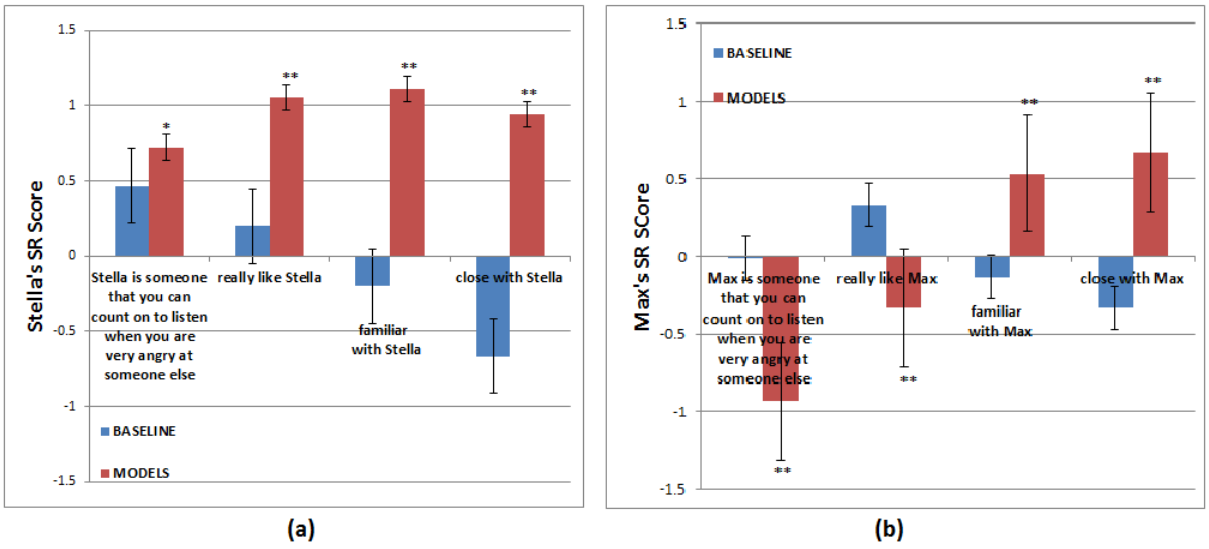


Figure 6.9: Stella's (a) and Max's (b) Social Relationship as Rated by Players

Player's Experience

Figure 6.10 illustrates the score of the players experience when playing the game. With the models implemented, the players feel more engaged and emotionally immersed in the game and the characters when playing with both NPCs. Players reported a significant increase in the engagement level with the game, with the models implemented ($p < 0.01$). A significant difference was also found in the players' emotional attachment to the game and the characters with the models implemented ($p < 0.01$). Moreover, the models enriched the game story, increased the players' curiosity how the game would progress ($p < 0.01$). From the short interview, most of the players preferred the NPCs with models. The models provide an interesting flow to forge the relationships with NPCs. However, some players suggested adding more variation to the interactions, as the current interaction is limited to speaking, giving items, and exploring Skyrim together.

On average, players spent around 0.92 hours (MAX = 3.2 hours, MIN = 0.25 hours) in game time and 62.78 hours (MAX = 114.2 hours, MIN = 11.75 hours) with Stella without and with models, respectively. Players spent 1.01 hours (MAX = 4.0 hours, MIN = 0.21 hours) in the game time and 74.74 hours (MAX = 219.5 hours, MIN = 16.17 hours) for Max with no

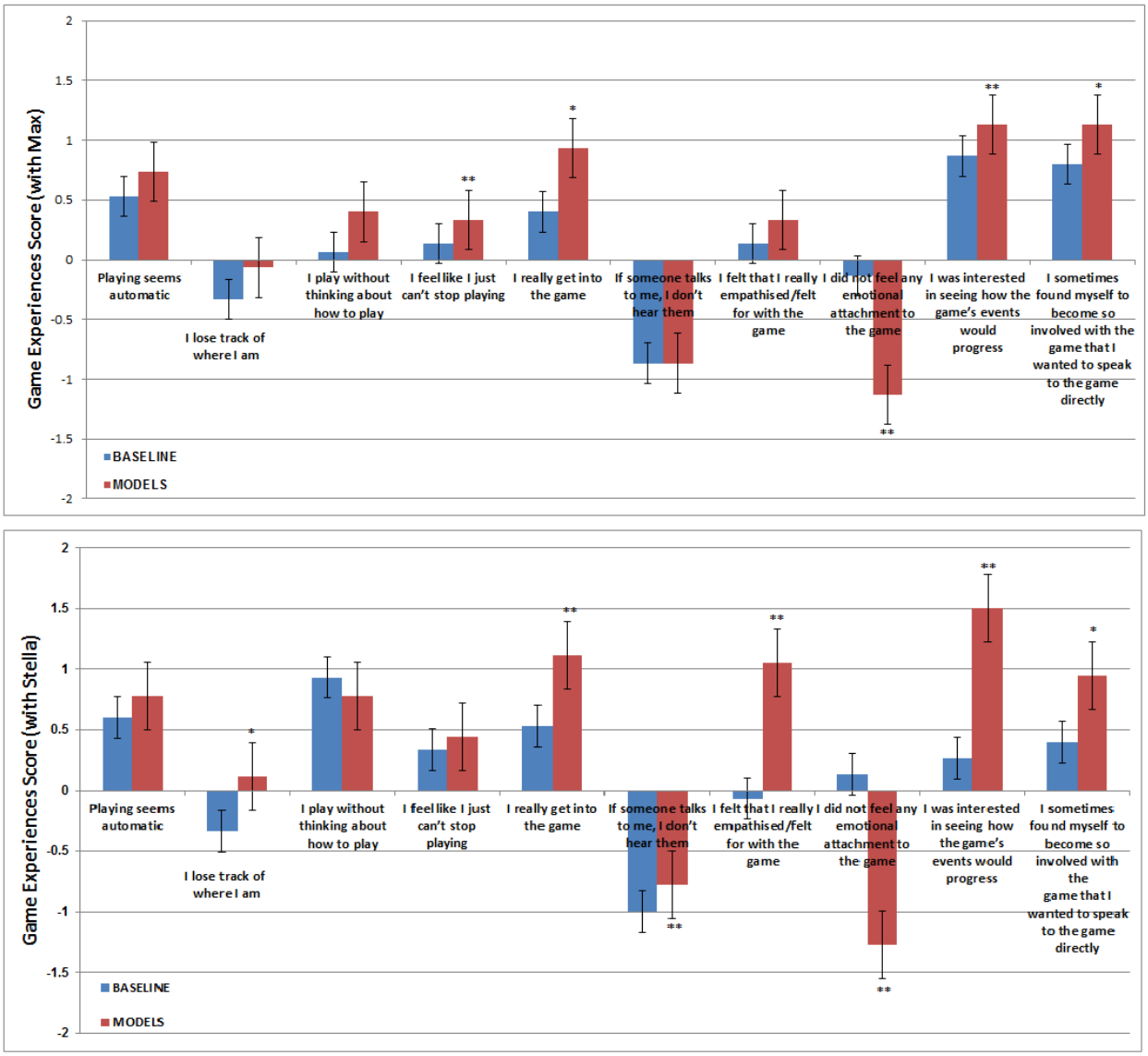


Figure 6.10: Players' Experiences with Max (top) and Stella (bottom)

models and with models implemented, respectively. Players who took a longer time to finish the quest spent their time mostly exploring the Skyrim world with the NPCs, while players with shorter time only focussed on forging a social relationship with the NPCs by giving them their most-liked items continuously. The favourite strategy to build relationships with the NPCs was giving favourable items to them, with exploring Skyrim together as the second choice of strategy. Building relationship with Max can be quite a nerve-racking experience, 40% of players realised that it would be easier to have a negative relationship and decided to go towards negative relation with Max although they already held a positive relationship with him as their game strategy to get The Sword.

Arousal and Valence Analysis

Similar to the Smile Game, automatic arousal and valence annotation were also performed from the interaction audio, in addition to manual annotation, in this study (see Figure 6.11 and Figure 6.12). There was a strong positive correlation between players' valence data with Max from automatic and manual annotation, $r = 0.76, p < 0.001$ and $r = 0.72, p < 0.001$, with baseline and models implemented respectively. In addition, there was also a strong positive correlation between players' arousal data with Max from automatic and manual annotation, $r = 0.62, p < 0.001$ and $r = 0.77, p < 0.001$, with baseline and models implemented respectively (see Figure 6.11).

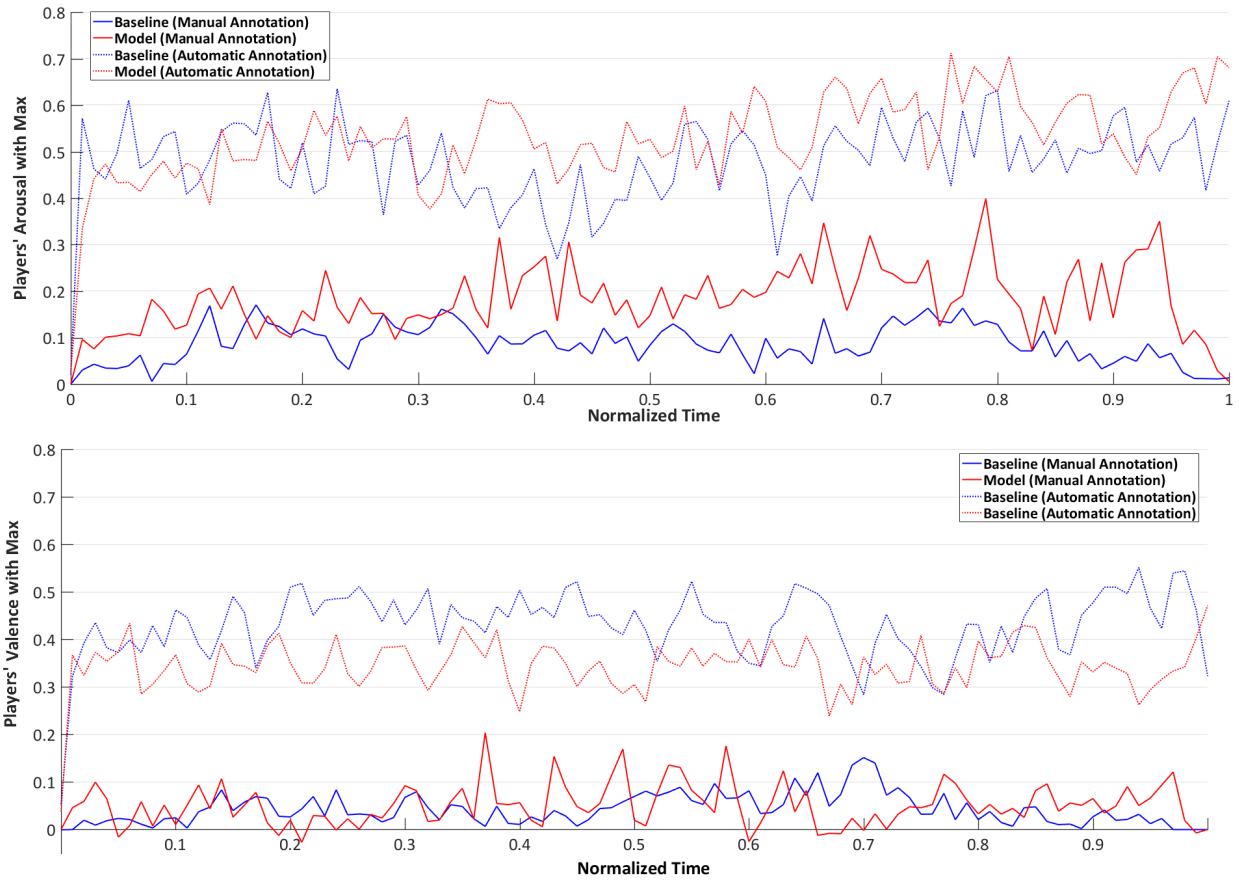


Figure 6.11: Players' Arousal (top), Valence (bottom) towards Max, where dotted line represents Automatic Annotation

Figure 6.12 demonstrates the players' arousal and valence when playing with Stella. There was also a strong positive correlation between players' valence data with Stella from automatic and manual annotation, $r = 0.72, p < 0.001$ and $r = 0.71, p < 0.001$, with baseline and

models implemented respectively. A strong positive correlation also found between players' arousal data with Stella from both annotations, $r = 0.68, p < 0.001$ and $r = 0.83, p < 0.001$, with baseline and models implemented respectively.

A Wilcoxon Signed-Ranks Test indicated that players' arousal was statically significantly higher with models implemented, $Z = 8.4, p < 0.00$ and $Z = 8.71, p < 0.00$, with Max and Stella respectively. In addition, players' valence was also significantly higher with models implemented, $Z = 3.1, p < 0.05$ and $Z = 7.75, p < 0.00$, with Max and Stella respectively. Negative valence also occurred in the interaction with ERISA empowered Max. Players' arousal and valence were significantly higher when playing with ERISA empowered Stella compared to Max, $Z = 3.75, p < 0.001$ and $Z = 6.47, p < 0.01$, respectively. On the contrary, there was no significant differences between players' arousal and valence when interacting with baseline Stella compare to Max, $Z = -1.67, p = 0.5$ and $Z = -1.65, p = 0.1$, respectively.

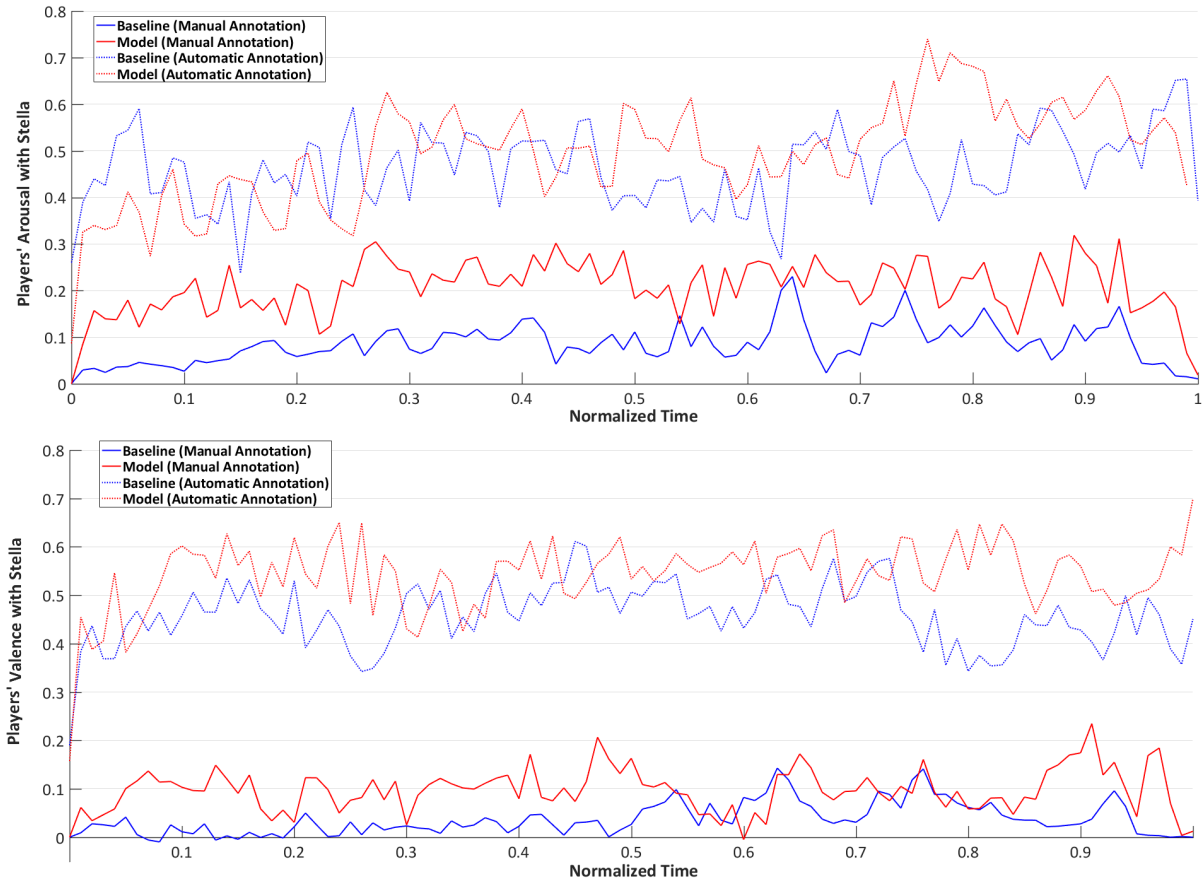


Figure 6.12: Players' Arousal (top), Valence (bottom) towards Stella, where dotted line represents Automatic Annotation

Figure 6.13, Figure 6.14 and Figure 6.15 show the boxplot for Player's Valence, Arousal and Interest Means Between Groups respectively for both automatic and manual annotation. Group 1 and 3 in the x-axis indicate the Stella Baseline and Max Baseline group respectively, Group 2 and 4 indicate Stella Enhanced and Max Enhanced group respectively. Player's valence towards Stella Model has the highest mean among the other groups, while Player's valence towards Max Baseline has the lowest mean among the others. Group 2 has a slightly higher median compared to Group 1. Around 90 percent of Group 2 is higher than the median of Group 1. Where there is no significant difference between medians of Group 3 and 4, but 75 percentiles of Group 4 is higher than 75 percentiles area of group 3.

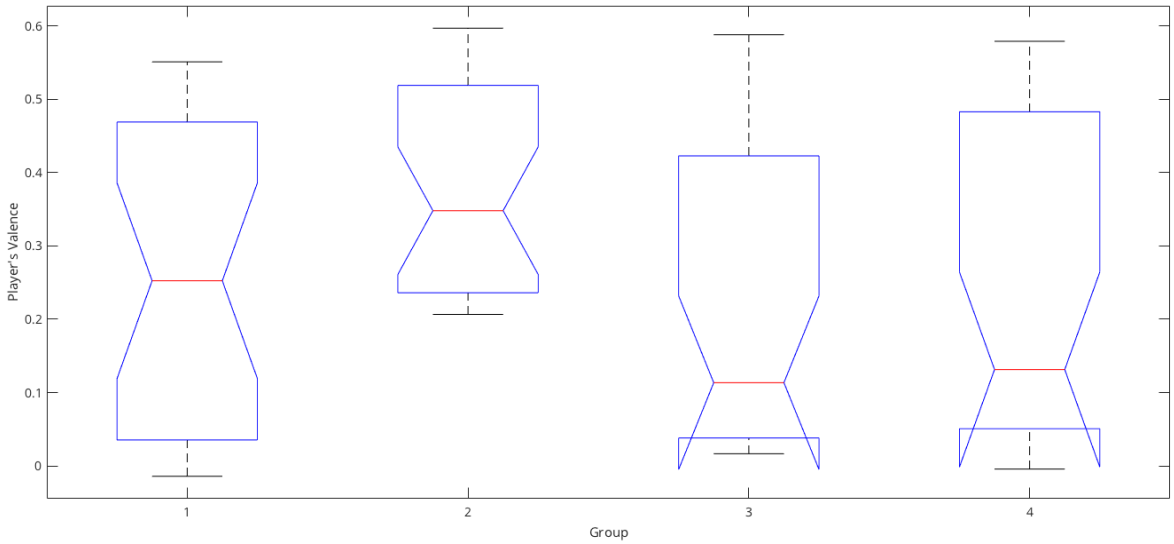


Figure 6.13: Player's Valence Means Between Groups

Player's Arousal towards Stella Enhanced has the highest mean among the other groups, while player's arousal towards Stella Baseline has the lowest mean among the others. The median of group 2 is significantly higher than group 1 as well as the median group of 4 is slightly higher than group 3. Both higher and lower 50 percentiles of group 2 and 4 are located in the upper 50 percentiles of group 1, and 3 indicated that lower 50 percentiles of group 2 and 3 are still higher than the median of the group 2 and 4. Lastly, there are no significant differences between the median of groups in the player's interest.

A MANOVA test was conducted to determine whether there are significant differences between the mean of a variable's mean between groups. The MANOVA examined between

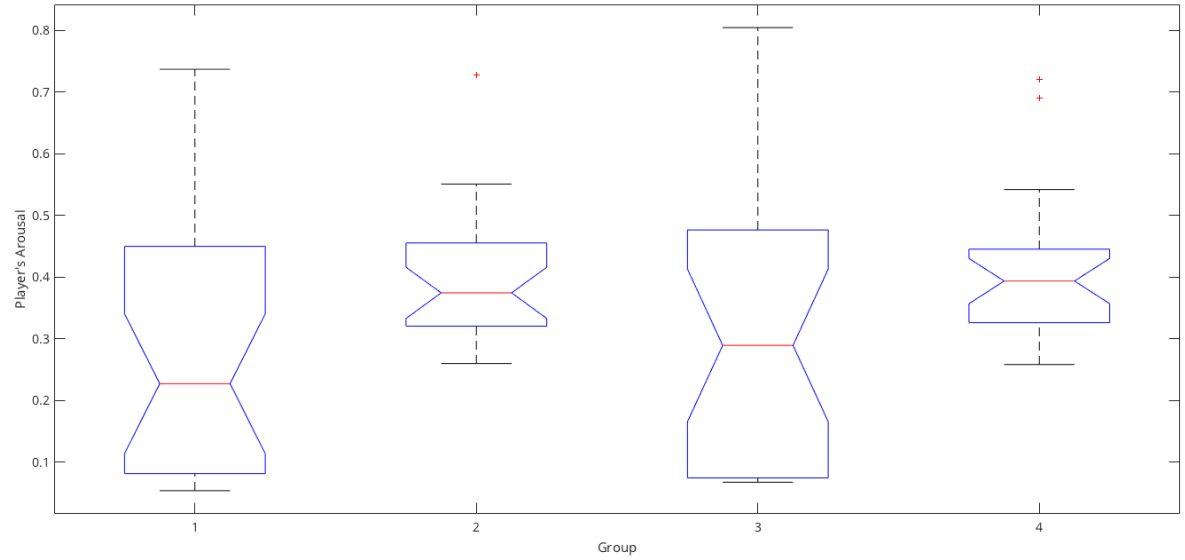


Figure 6.14: Player's Arousal Means Between Groups

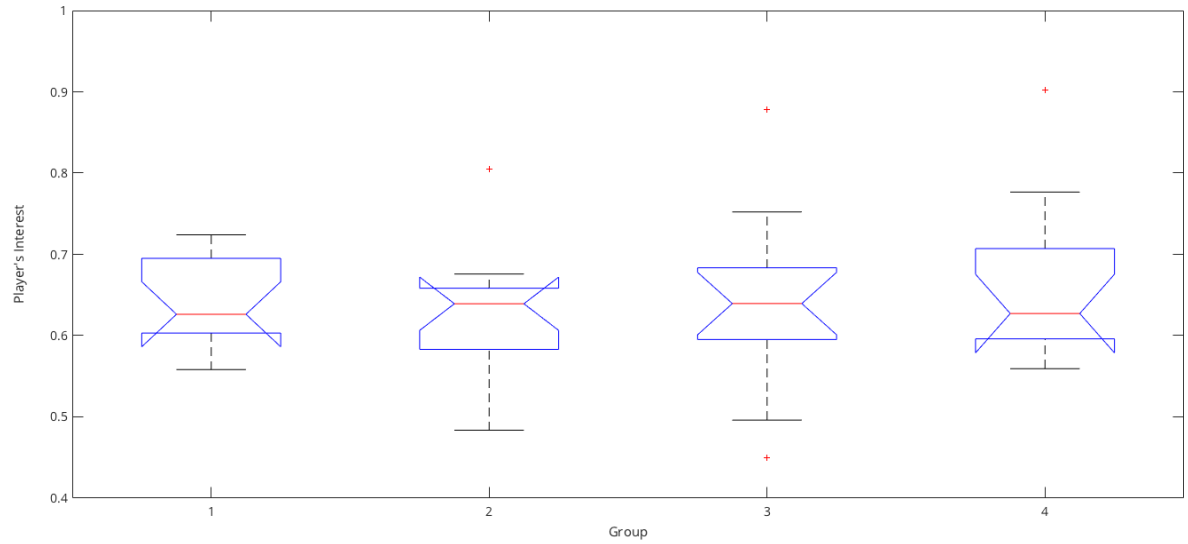


Figure 6.15: Player's Interest Means Between Groups

groups of Max Baseline, Max Enhanced, Stella Baseline and Stella Enhanced, with variables of manually annotated valance and arousal as well as automatically annotated valence, arousal and interest. The results indicated that the means lie on the same plane in 3-dimensional space but not on the same line ($d = 2, p < 0.001$) with Wilk's lambda value of $0.0109, 0.1608 and 0.9631$. To determine which two of groups differ from the other, a canonical analysis was conducted. Figure 6.16 illustrates the separation between groups using a canonical analysis. The figure shown that two groups that contain mean differ with the other are both enhanced models: Stella Enhanced and Max Enhanced. The Stella Enhanced

group are mostly spreading in the first quadrant of the axes, while the Max Enhanced group are mostly located in the third quadrant of the axes. The Baseline models (i.e. Max Baseline and Stella Baseline) are both located in the fourth quadrant of the axes.

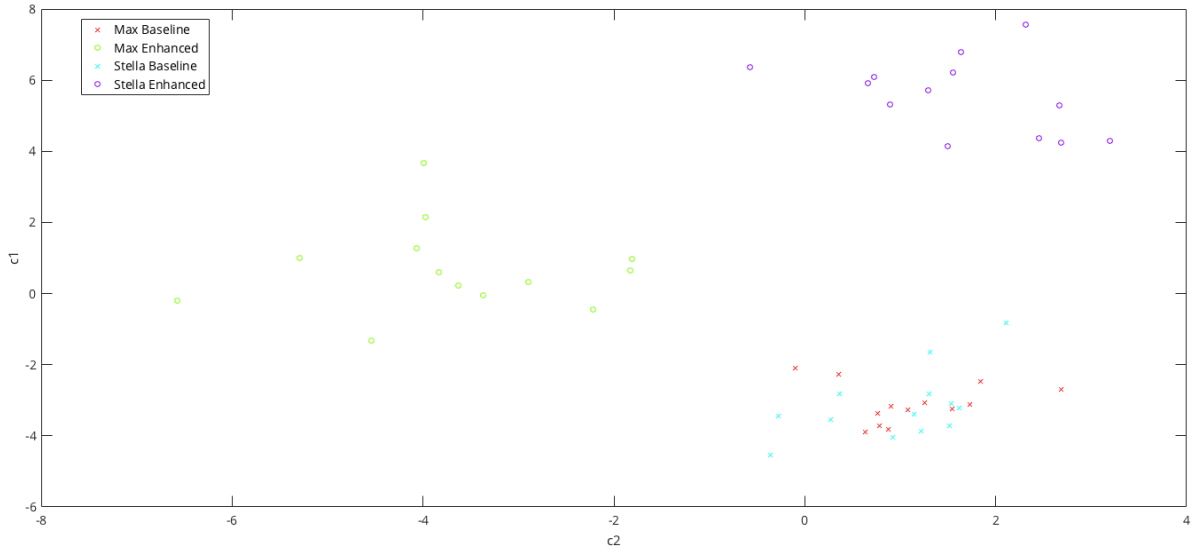


Figure 6.16: Canonical Analysis

6.4 Conclusion

The study clearly provides the empirical findings of a more natural and believable interaction between a player and game companions leading to a simple social relationship between them. With the framework, the game companions were capable of perceiving the player's affective signals from their facial expressions as well as to exhibit social signals from the facial expressions, gestures and verbal cues (**RQ1**). The personality of NPCs with models implemented is consistent with the personality that we set to both NPCs. The results show that the players evidently perceived the game companions' personality in accordance with the one that the game designer assigned to them. The game companions were able to display emotions appropriately, providing an enhanced feeling of immersion in games to the players. Moreover, the ability to forge relationships naturally with the game companions resulted in an emotional attachment towards the game companions. Moreover, players were emotionally involved and attached to the characters and the game. Social relationship models also proved

to provide a more realistic manner to interact with game companions and forge relationships with them (**RQ2**). The players were able to tell what relationship a player has with a game companion in a game immediately from their attitudes shown by their facial expressions, gestures and verbal cues (**RQ2** and **RQ3**)

NPC personality also influences the gameplay, as the characters with the Neuroticism trait were harder to defeat thus requiring a personality-specific strategy to win. Social relationship models can be adjusted to enrich the game story as well. For future research direction with these models, more variation for interactions with the NPCs can be extended. Interaction such as accomplishing a quest together and more conversational topics can enhance the user experience when playing the game. For future research direction with these models, more variation for interactions with the NPCs can be added. Interaction such as accomplishing a quest together and more conversational topics can enhance the user experience when playing the game. Moreover, there is also a significant change in the social relationship score shown in the results of models implemented with both NPCs. The players were able to tell what relationship a player has with a game companion in a game immediately from their attitudes shown by their facial expressions, gestures and verbal cues (**RQ2** and **RQ3**). Furthermore, the models provide a new experience in playing games to the player, particularly in relation to the emotional attachment to the NPCs and how they build a social relationship with them. The models provide a more realistic manner to interact with NPCs and forge relationships with them. In addition to the conventional game inputs (i.e. joystick, mouse, and keyboard), the models and framework allow the player to interact with the game characters with a camera by using their facial expressions (**RQ6**). For future research direction with these models, more variation for interactions with the NPCs can be added. Interaction such as accomplishing a quest together and more conversational topics can enhance the user experience when playing the game.

[This page intentionally left blank]

Chapter 7

Game Character Design Workshop

7.1 Introduction

To further explore the potential of the ERiSA models and framework for game development, a character design workshop was conducted with game designers as participants. A focused design workshop was conducted to explore the potential of the ERiSA framework and models. Section 7.3 illustrates the workshop setup. The workshop took place in our university for two hours and attended by two experts in the game and/or character designers. The aim was to investigate how game designers would use the models and framework to design new game companions. Moreover, the workshop allows us to investigate and learn from the participants, who are game designer and artist, about their perspective and experiences when they were using the tool provided. A simple tool to simulate the proposed models was created to be used in the workshop. The overview, settings and configurations of the tools are described in section 7.2. Moreover, section 7.4 shows the results of the workshop. From the discussion, we can conclude that the models and framework can be used to create a game that has emotions and social relationships elements. In a particular game such as crowd simulation, life simulation, character-driven and serious games. In addition, with the experts in their field, the ERiSA framework and models have the potential to be a tool for evaluation or simulation in a real-world scenario.

7.2 The ERiSA Tool

ERiSA Tool is a simple tool to simulate the computational models of emotions, personality and social relationships proposed in this thesis (see Section 3.3). The user will be able to see in real time the effect of a particular event to the virtual human's emotions and social relationships and vice versa by adjusting a number of parameters provided (see Figure 7.1). The user will also see the association between the virtual human's personality and the events simulated. The tool will provide a particular answer based on the parameters of the models and events set by the user.

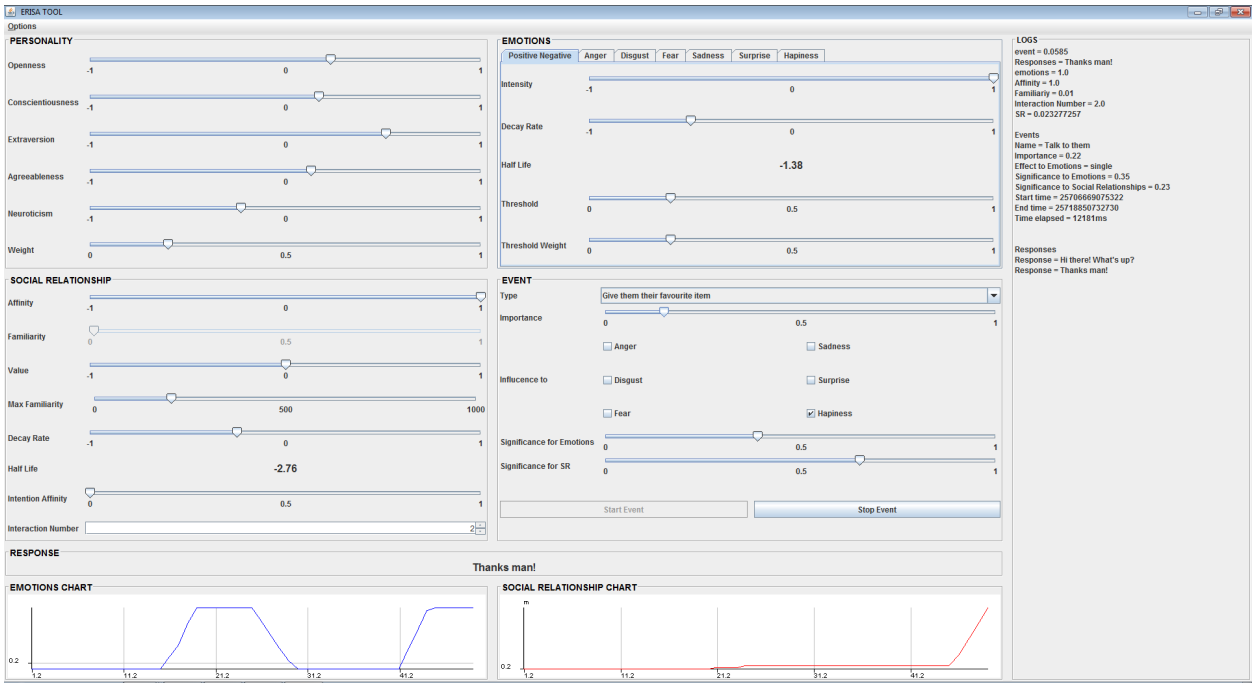


Figure 7.1: ERISA Tool for Character Design Simulation

The tool's environment is divided into three areas: upper-left, bottom-left and right area. All the computational models' parameters are presented in the upper-left area providing the playground for the user. The tool provides a simple and advanced mode to adjust the parameters. In the advanced mode, the user will be given all the parameters to be played around, while in the simple mode the user only has access to some important parameters. Parameters such as Personality Weight, Emotion Decay Rate, Emotion Half-Life, Emotion Threshold, Emotion Threshold Weight, Social Relationship Decay Rate, Social Relationship Half-Life, Affinity and Familiarity values, Event influence to emotions are not visible to the

user and set to their default value. Over time the user changes any of the parameter's value, they can see the effect of the changes to the virtual human's emotions and social relationship in a real-time through two charts displayed in the bottom-left (left side is emotions chart, and the other side is social relationship chart). Moreover, all the logs are displayed in the view panel in the right area of the tool. The user can save the parameters values and the logs and load them any-time. The parameters' value can also be randomised using the menu located in the Options menu. Finally, in the future development, the user will be able to link the tool to The Skyrim Creation Tool Kit, so the user can use the characters in the Skyrim to perform the actions in the Skyrim world by using Skyrim Creation Toolkit to bridge the framework, tool and the Skyrim Character in the Skyrim game world.

7.3 Experimental Setup

A two hours workshop was attended by two game and/or character designers with three to four years of experience in the designing game or the character for a game. One of the participants was a game designer and programmer with four years of experience in the making Role Playing Game (RPG) (Male; Age: 27, Caucasian). The other one was a 2D game characters artist with more than three years of experience in the field (Female; Age: 24, Asian). Both participants designed and created a number of games or game characters but never published them to the public. The main aim of the workshop was to explore the potential of the models and framework to design characters for games from the point of view of a character or game designer.

In the workshop, the researcher introduced the framework and models to the participants and provided some examples of what the game designers could do with the models and framework. Next, the participants were asked to design three characters: one based on a real person that they know in real life; second, based on the same person but now with the opposite personality; and the third character could be any personality of the participant's choice. A tool was designed to simulate the ERiSA models. With the tool, the participants

were able to design a character by adjusting parameters of emotions, personality and social relationship of ERiSA models. The parameters can be seen in Table 3.2. The participants were also asked to simulate an interaction between them and the character with the tool, and the tool provided an answer based on those parameters and event evoked. Figure 7.1 illustrates the features provided by the tool. The tool will automatically save the profile of the character designed and all the interactions logs (i.e. the events induced to the character) when the tool is closed. Alternatively, the logs and character profile can be manually saved with a menu under Options. Also, the character profile can be loaded to get all the variables assigned to the character previously. All the variables and logs are stored in a text file format, where it also can be manually read with any text editors. Finally, the workshop was closed with a discussion regarding the characters designed by the participants, the potential of the ERiSA models and framework to be implemented in game design and game scenarios based on the characters designed by the participants. The discussion method was a semi-structured interview with four categories of question: The participant's Basic Profile (PRO), the participant's opinion about the tools and framework (OPI), game design with the tool (GDN), character design with the tool (CDN). All the discussions were audio recorded to be further analysed with the coding system. All the characters design data parameters (e.g. Personality, Emotions, Social Relationship), as well as the interaction data, were also logged. Table 7.1 demonstrates all the interaction data collected from one of the participants.

7.4 Results

7.4.1 Character Design Task

In the first design activity, both participants designed the characters based on their own perception towards the character in the real world. Participant 1 designed characters that represented Donald Trump and participant 2 choose her best friend as her model. From the participant 1 perspective, a character that represents Donald Trump has high Extraversion, Agreeableness and Neuroticism trait value. The participant also describes

that the character represents Donald Trump should also has a low value of Conscientiousness and Openness ($\mathcal{P}_{O,C,E,A,N} = \{0.26, -0.21, 0.71, 0.71, 0.43\}$). In the participant 2 mind, a character that represents her best friend should have high Extroversion and Agreeableness traits also participant 2 describes that her best friend has a low value of Neuroticism ($\mathcal{P}_{O,C,E,A,N} = \{0.00, 0.00, 0.87, 0.68, -0.17\}$).

Table 7.1: Interaction with Characters at The Workshop

Character	**Event**	**Start**	**End**	$\mathcal{E}_{v+,t}$	$\mathcal{M}_{v,t}$	$\mathcal{S}_{i,j}$
I	I	1	79	0.59	0.61	0.16
	II	92	180	0.50	0.25	0.22
	III	188	434	-0.47	0.37	0.66
	IV	460	611	-0.65	0.46	0.95
II	I	1	198	0.05	0.14	0.32
	II	206	401	-0.92	0.62	0.33
	III	407	489	0.25	0.39	0.38
	IV	500	740	0.56	0.26	0.45
	V	745	796	-0.13	0.36	0.51
	VI	815	984	-0.60	0.45	0.25
	VII	998	1031	-0.80	0.70	0.50
	VIII	1042	1081	0.40	0.50	0.30
III	I	1	72	0.60	0.54	0.40
	II	82	155	0.46	0.64	0.53
	III	172	204	0.10	0.20	0.20
	IV	224	248	-0.42	0.88	0.17
	V	352	454	0.63	0.26	0.14
	VI	460	506	-0.35	0.41	0.19

In the second design activity, the participants designed characters that have opposite characteristics with the characters they designed in the first activity. From the participant 1

perspective, a persona that has opposite personality with his first character should have a high value of Openness with a low value of Agreeableness and Neuroticism trait ($\mathcal{P}_{O,C,E,A,N} = \{0.48, 0, 0.13, -0.37, -0.37\}$). While participant 2 think that a character that represents the opposite personality from her best friend would be like a very Neurotic person with a low value of Extraversion, Openness and Agreeableness ($\mathcal{P}_{O,C,E,A,N} = \{-0.24, 0.00, -0.43, -0.57, 0.62\}$).

Finally, in the last design activity, both participants set some parameters to an extreme value. Participant 1 designed a character that has extremely high Openness, Extraversion and Neuroticism ($\mathcal{P}_{O,C,E,A,N} = \{1.00, 0.10, 1.00, -0.26, 0.93\}$). The participant also explained that the character he designed has a high intention for a new relationship. Participant 2 designed a character that has extremely high Neuroticism and reverse Extraversion value ($\mathcal{P}_{O,C,E,A,N} = \{0.00, 0.00, -1.00, 0.00, 1.00\}$). In her mind, this character should behave like an obnoxious kid.

7.4.2 Interaction Simulations with the Characters

In this workshop, the participants were also asked to interact with the characters that they have designed with the tool provided. There were events provided in the tools simulating positive and negative events: Talk to the character, Give the character their favourite item, Give the character their undesirable item, Praise the character and criticise the character. The logs of the interaction were recorded to be further analysed. All participants started the interaction with the characters with a positive event. The participants then started to explore the effect of both positive and negative events to their designed character with various parameters. Figure 7.2 illustrates the results of the simulations of the interactions with the characters that one of the participants designed.

Figure 7.2 (a) demonstrates the interaction with the first ($\mathcal{P}_{O,C,E,A,N} = \{0.26, -0.21, 0.71, 0.71, 0.43\}$), (b) with the second ($\mathcal{P}_{O,C,E,A,N} = \{0.48, 0, 0.13, -0.37, -0.37\}$) and (c) with the third character ($\mathcal{P}_{O,C,E,A,N} = \{1.00, 0.10, 1.00, -0.26, 0.93\}$). The other parameters set to the first character were: $\mathcal{Y} = 56$, $\mathcal{N}_{i,j,t} = 0.6$, decay rate for social relationships models $r = -0.25$, $\mu = ln(2)/r$, emotion decay rate $r = -0.5$, $\mu = ln(2)/r$, negative emotions' basic

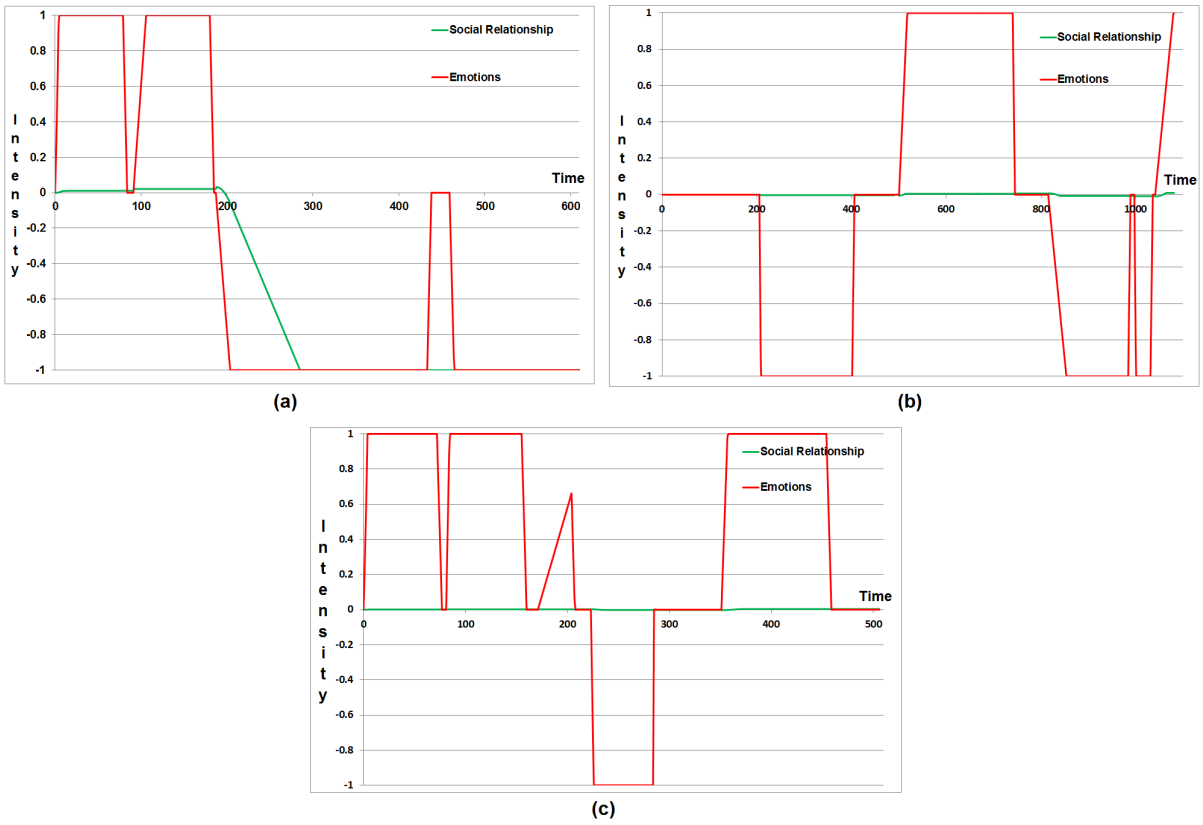


Figure 7.2: ERISA Tool for Character Design Simulation

threshold $\theta_{v-} = -0.2$, positive emotions' basic threshold $\theta_{v+} = 0.2$, $\mathcal{W}_p = 0.2$. Moreover, the other parameters set to the second character were: $\mathcal{Y} = 495$, $\mathcal{N}_{i,j,t} = 0.6$, decay rate for social relationships models $r = -0.25$, $\mu = ln(2)/r$, emotion decay rate $r = -0.5$, $\mu = ln(2)/r$, negative emotions' basic threshold $\theta_{v-} = -0.2$, positive emotions' basic threshold $\theta_{v+} = 0.2$, $\mathcal{W}_p = 0.2$. Finally, the parameters set to the third character were: $\mathcal{Y} = 430$, $\mathcal{N}_{i,j,t} = 0.25$, decay rate for social relationships models $r = -0.19$, $\mu = ln(2)/r$, emotion decay rate $r = -0.5$, $\mu = ln(2)/r$, negative emotions' basic threshold $\theta_{v-} = -0.2$, positive emotions' basic threshold $\theta_{v+} = 0.2$, $\mathcal{W}_p = 0.2$. A participant mentioned that generally game designers tend to exaggerate the character's personality so the player can distinguish the character's personality. This motivate the participant to exaggerate the third character's personality.

There was a total of 4 interactions in 611 s with the first character, eight interactions in 1081 s with the second character and six interactions in 506 s with the third character. Table 7.1 illustrates the overview of all events simulated to all three characters. Compare to the interaction with the other characters, the first interaction was resulting in an extremely

negative relationship between the character and the player. This was because the third and fourth events had a high impact on both the event itself (-0.47 and -0.65) and the impact to the social relationships (0.66 and 0.95) and lasted for more than half (65%) of the interaction. In addition, compared to the other character, the first character has a low number of maximum familiarity, $\mathcal{Y} = 56$ compare than the other ($\mathcal{Y} \geq 430$), resulting in a fast effect on the social relationship. We can already see the relationship between them in only a couple of interactions. In contrast, the social relationships with the second and third characters were relatively hard to see, as the value of the social relationships were only between -0.1 to 0.1 ($-0.1 \geq \mathcal{R}_{i,j,t} \leq 0.1$). This is due to the impact of the events on the value of the social relationship were set not too significant in the second and third characters. Moreover, the value of the maximum familiarity set to both characters was also too high to be achieved.

7.4.3 FGD: The Potential of The Models and Framework

A semi-structured discussion was conducted to explore the potential of the proposed models and framework. The Coding system was applied in the analysis process. Table 7.2 demonstrates all the questions the group asked during the discussion. There were four question's categories: questions about the participants' profile, participant's opinions, game design using the framework and tool and designing a character using the framework and tool.

Table 7.2: Questions

Code	**Category**	**Detail**
PRO	Profile	Participant's profile and demography
OPI	Opinion	Participant's opinion about the framework and tool
GDN	Game Design	Game design using the framework and tool
CDN	Character Design	Character design using the framework and tool

Table 7.3 shows the coding system for the participants' answers during the discussion. In summary, in the question group of PRO, we coded P1 as the profile of Participant 1 and P2 as the profile of Participant 2. The questions group of OPI there were two categories of the answer: PROCESS and FEATURES. Also, the questions group of GDN had four categories

Table 7.3: Discussion's Summary

Question	Answer's Code	Detail
PRO	P1	Male, Game Designer & Developer, 4 years experience
	P2	Female, Artist, 3 years experience
OPI	PROCESS	Game and character design and development process
	FEATURES	The framework and tool features for game and character design and development process
GDN	STORY	Game design focused in the story aspect
	CHARACTER	Game design focused in the character aspect
	SIMULATION	Game design focused in simulation
	SCENARIO	Game design focused in a particular scenario, be they real or imaginary
CDN	EMOTION	Emotion aspect in the character design
	SR	Social relationship in the character design
	PERSONALITY	Personality aspect in the character design

of the answer: STORY, CHARACTER, SIMULATION and SCENARIO. Lastly, the group of CDN had three categories of the answer: EMOTION, SR and PERSONALITY.

Both participants argued that the tool and framework would offer a lot of possibility for character and game designer. Participant 1 argued that the tools and framework would "... speed up the game design process." (PROCESS) and Participant 2 commented that "... make it easy to see the character design implication" (PROCESS and FEATURES). Both participants mentioned that in the conventional ways to design a character or game, the process of designing and testing the character is done in a repetitive manner hence it requires a lot of time. Participant 1 argues that " The tool makes it much faster" (PROCESS). However, both participants argued that the tool and framework are probably not suitable in the entertainment industry other than the game industry. Participant 1 quoted "... but I guess people in the movie industry still prefer real human to test it ... have real emotional responses." (FEATURES and EMOTION).

Some potential games can be designed with the ERiSA models, and framework concluded from the discussion during the workshop. The games can be categorised into five groups of applications. The framework and models can be potentially to be implemented into characters in games such as crowd simulation, life simulation, character-driven and serious games. Furthermore, participants also mentioned the application of the framework and

models to evaluate/simulate a real scenario in the real world with guidance from the experts in their field depending on the scenario. Below are the categories:

1. **Crowd Simulation** Both participants agreed that a crowd simulation game such as GTA could be a good implementation for the models and framework. Imagine all the characters in the game have their own unique personality and it requires a different strategy to get along with the characters and finish the events depending on their personality and the nature of the social relationship between the NPCs and the players (SIMULATION, CHARACTER, EMOTION, SR, PERSONALITY). The designers will be able to design the level with changing the characters parameters in addition to the others aspects of the game. For example: to make a character to be difficult to start a new relationship with the player, the designer can create a character with high Neuroticism or low Extraversion and set its maximum familiarity and intention to have a new relationship to a high value. The player can also use their facial expressions as one of the modal to interact with the characters in the game. The participants believe that such a game will provide a new experience to play a game.

2. **Life Simulation** The other game that can fit for the models and frameworks for its characters is the life simulation games such as the Sims (SIMULATION). With the models, the player can design not only the character's outer outfit or appearances but also their personality (PERSONALITY). With the model, the player will be able to create a character that represents the player in the virtual world. With the framework, the player also can interact with other characters with their facial expressions (FEATURES), which the participants believe will provide a new experience toward the game. A participant suggested that we can make a game that the goal is to make Trump as angry as possible. The angrier he is, the higher score the player will get. Eventually, if the player can push him hard enough, he will get a heart attack, and the player wins the stage. The model can simulate Trump's personality, emotions and social relationship parameters and the framework will provide the ability for a player to naturally interact with Trump (e.g. using facial expressions or automatic

speech recognition and natural speech processing) (FEATURE, CHARACTER, SCENARIO, EMOTION, SR, PERSONALITY). The Participant quoted "If you push it hard enough then he would have a heart attack and die." The other participant suggested similar games, but instead of making angry with the character, the player should attract the characters by using different strategies based on the character's parameters designed using the models. The other participant mentioned "... imagine him as your dream boyfriend, and you have to make him to like you." (CHARACTER, STORY, SCENARIO)

3. **Character Driven** The participants also mentioned that games with character driven or focus could be fitted for the models and framework (e.g. The Last of Us[1]). Such games generally rich in the story to shape the character, so the players also enjoyed the story as the player is progressing through the game levels (STORY, CHARACTER). Participant 1 quoted "The tool is suitable to design a game where a story is important." Participant 1 also commented "So, it is basically allow you to like... simulate what it would look like"". Based on the discussion, the implementation of the models and framework will augment the richness of the story through its characters. Moreover, the story can be designed depending on the player's actions and choices to the NPCs. These multiple endings will thus provide a new experience to the players (STORY, SCENARIO).

4. **Serious Games** There is a potential to design a serious game with this framework and models. The models provide the capability to design a realistic but limited simulation on how people should behave based on a specific event designed to a specific scenario (SCENARIO). The framework gives the capability to read player's facial expressions and mapped them to the games (FEATURES). Serious games for autism players was one of the game example discussed in the workshop. With the supervision of the experts, the game will be a potential tool to help the player to understand and display the appropriate emotions. The player's scores or points, for example, will be reduced if they can not identify the correct emotions displayed or are not able to reproduce the

[1]http://www.thelastofus.playstation.com/index.html

same emotions acted by the game characters (FEATURES, SCENARIO).

5. **Evaluation Tools** Finally, Participant 1 argued that the framework and models could be potentially utilised as evaluations tools in real scenarios with the supervision of the experts (SCENARIO). We can implement the models and framework to a virtual character and test it with the real human to get the overview of the real emotions responses (EMOTION). A more interesting way to test the framework and models is to test it with a real human and see whether with the same event stimulated to both real and virtual human, would produce identical behaviours. Participant 1 also suggested that "it would be nice if you could load up two characters simultaneously and have it side by side comparison, to... to... to be able to see the differences reactions to different inputs." (FEATURE). Furthermore, Participant 2 suggested that more details dialogue in the conversation provided in the tool would make more realistic to simulate a particular scenario (FEATURE).

7.5 Conclusion

During the design workshop, the participants concluded that the framework and models provide the ability to enrich the game characters by shaping them with the variable of emotions, personality and social relationship (**RQ1**, **RQ2** and **RQ5**). Current conventional game design still using a static manner to the character's parameters. However, with this models and framework, the designers are now able to dynamically design the character emotions, personality and social relationship to the game. The models and framework offer many possibilities for character designers to make it easy to see what the design implication are to games directly and the characters in real time and thus speed up the design process. The participants thought it would be perfect if the tool can load up two characters simultaneously and see the difference between the character with the same event evoked to the characters. A more detail conversation would also provide more realistic to simulate some scenarios.

Chapter 8

Discussion

This thesis presents computational models of emotions, personality and social relationship for virtual human as well as a modular framework for a social-affect-aware game that focuses on the social interaction between player and virtual human. The models allow the game designer to design a character for games with rich personality, able to convey emotions and forge a social relationship with the players by manipulating the value of the variables in the models and framework. The framework provides an opportunity to provide a new interaction form resulting in enhancing the player's experiences when playing games by enabling the player to use their facial expressions and simple words to interact with the virtual human. The models and framework were implemented in several game scenarios with simple social interaction through affective loop experiences. This thesis also explores virtual human and human interaction in the wild. Several data were collected from the players' interaction to craft the interaction rules for the virtual human from the studies. The results of the studies indicate that the players were able to clearly perceive the virtual human's personality in accordance to the one set to them. The players were also able to forge social relationships with them and perceive the relationship they have with them immediately when it changed. There were significant improvements in the players' experiences in the games with the models and framework implemented. Particularly the emotional involvement with the virtual human resulted in a positive enhancement to the game immersion.

8.1 Research Contributions

8.1.1 The Models and Framework

This thesis proposes a modular and integrated framework for social and emotional virtual human, complete with the ability to develop simple social relations with the player over time, by extending an existing framework for building emotion-oriented systems, SEMAINE (Schröder 2010) and implementing it in a game environment. In addition, this thesis also introduces integrated computational models of emotions, personality and social relationships for a formulation of ruled based action selection to enrich the interaction between players and the virtual human (shown in Chapter 3). The models were carefully shaped from the basis of existing theories, from the field of psychology and social sciences. These models and framework are aimed to fill a research gap in providing a social-affect aware framework to develop a game.

The implementation and evaluation of the framework and computational models are demonstrated in Chapter 4, 5, and 6. Those chapters present the process of establishment of natural and believable interaction between a player and game companion to create affective loop experience that is informed by the process of encoding and decoding of social signals exchanged between the player and game companion. Moreover, Chapter 7 illustrates the potential of the proposed framework and computational models to be implemented in the game industry. With the ERiSA models, the game designer will have an opportunity to create a believable game character by providing a rich characteristic from the personality, induced emotions and provide the ability to forge relationships with the players. The ERiSA models were formed from an existing theory from the field of psychology and social sciences. The primary focus of the models is to manipulate three behavioural variables that are mostly induced to the virtual human to increase their believability. Some excellent works have been done in this area; however, only a few of them take into consideration the interaction between those components. Most games are also still implementing static emotions and relationships where the effect of the emotions and relationships between the players and NPCs are still

using a "click-based" system. This method results in the virtual human generating the same static behaviours over and over again thus allowing the players to predict the behaviours and affecting negatively to the players' experiences when playing the game and interacting with the NPCs (Ochs et al. 2009). Nevertheless, in the real world, emotions, personality and social relationships would have influences on each other (see Figure 3.1).

The framework and computational models are asynchronously communicating with each other with ActiveMQ. Hence, it provides the possibility to plug and play components to the framework without causing any issues to the system, as long as they are using the same topic in sending and receiving messages. The second game scenario, The Skyrim game demonstrates the ability to plug and play components of the framework. In this scenario, several components of the framework were used in addition to additional components of the Skyrim implemented to the framework (e.g. the part of dialogue manager and the output system). Moreover, the framework consists of Memories that are stored in XML based files as the configuration files to the virtual human. The files consist of a static memory file is generally never be updated apart from the initial design of for game balancing and a dynamic memory file where it will frequently be updated (see Appendix C).

This thesis argues that a virtual human equipped with the computational models and framework will increase the player's experiences when playing the game and provide the empirical evidence gathered from some studies done in two game and interaction scenarios to validate this argument. Moreover, based on the discussion on the workshop design attended by game designers who have been working for three to four years that the models and framework offer many possibilities for game character designers to make it easy to see what the design implication are to games directly and the characters in real time and thus speed up the design process. The workshop's participants also mentioned that the ERiSA framework and models could be implemented to design a game that involves emotions and social relationships elements. Mainly games such as crowd simulation, life simulation, character-driven and serious games (see section 7 for the details). The studies (i.e. The Smile Game and The Skyrim Game) demonstrate that the game designers could incorporate the framework and models with the game design, particularly in the game levelling and story. In both studies,

the variable personality was used to adjust the difficulty of the game. In The Smile Game, nice personality virtual human resulted in easier opponent to defeat and vice versa. In The Skyrim Game, nasty personality virtual human provides more challenges to get info from them and vice versa. There are of course some limitations of the models and frameworks. The face recognition and face expression recognition system in the framework only capable of recognising and processing one person at one time. The dialogue system was also not able to fully understand all the user's utterance as the system is only capable of spotting some words. Finally, not all facial action units or emotions can be synthesised due to the limitation of the rule-based system in the Action Manager component.

8.1.2 Data Collections for Interaction Rules in Games

Chapter 4 and 5 present the design and protocol to collect and craft the interaction rules for the virtual NPCs in the game environment to support the creation of affective loop experiences. Collecting information of interaction between human-human and human-virtual human in a public setting is paramount to create a reliable virtual human. This is due to the uncertainty of the interaction in the public space setting. The Smile Game studies conducted in the Public and Semi-Public space setting demonstrate proved this notion. This thesis collects more than three thousand annotations as the ingredient to craft the interaction rules (both conversational and game) for the virtual human. The data were collected from The SEMAINE Database Corpus, recording of Human-Human interaction and Human-Agent interaction in some *The Smile Game* studies setting. Ideally, collecting data of social interaction is favoured to craft the behaviour rules for virtual human, however, in reality, we are limited in time and resources to collect and process the data. On the other hand, collecting insufficient data might lead to improper behaviours shown by the virtual human as there are not enough data to be learnt by the virtual human. Therefore, the ideal condition to get adequate behaviour for the virtual human is to balance the data with the dialogue manager technique. To do so, first, we need to understand what would people generally ask and do during social interaction with a virtual human in the public space setting. The analogy is

like learning a new language to communicate with a foreigner. The first step is to learn words and sentences that the foreigner use in a particular situation (e.g. asking the price, asking the direction). Moreover, we can learn how normally the foreigner behaves during the interaction by copying them. In such manner, we need to simulate a real interaction between people and virtual human in the public space. We could use the *Wizard of Oz* technique to simulate a system that is not fully functioning yet to collect data as well as to evaluate the system (Green & Wei-Haas 1985*b*). There were two categories of interaction in this thesis, namely interaction during social conversation and interaction during the game (playing). Therefore, this thesis demonstrates two major annotation schemes, Dialogue Acts annotation (Austin 1975) and FACS Ekman et al. (2002*a*). The Dialogue Acts annotation aims to annotate verbal (e.g. interlocutor's utterances) and FACS annotation aims to annotate the user's behaviours when playing the game (e.g. when attacking, defending, or others).

In total, forty-six adult participants participated in the data collections with a total of 103 interactions. Ten of them participated in the Human-Human interaction data collection, while the other 36 were interacted with the virtual human and being video and audio recorded. Chapter 4 provides details of the data extraction, and annotation from an existing interaction corpus, The SEMAINE Database Corpus and Chapter 5 illustrates an exhaustive detail on the methods and results of the data collected from the Human-Human interaction and Human-Agent interaction in *The Smile Game* studies setting to produce behaviours specific to our scenarios. As an initial step to craft the interaction rules between the virtual human and the players, the transcripts of the conversations between a human and an operator controlled virtual agent from The SEMAINE Database Corpus McKeown et al. (2012) were annotated. The transcripts were annotated with Dialogue Acts annotation Austin (1975). There was a total of 2100 annotations from 40 transcripts, composed in 54 dialogue acts. The adjacency pair approach was adopted to determine pairs of the utterances. For example, a dialogue act of ask-wh-question-aboutagent can be paired with answer-wh-question-aboutagent. From there, a set of dialogue rules can be crafted. It is also possible to add valence to the answer depends on the emotions of the virtual human by annotating the valence at the end of the dialogue acts. For example: Suppose a default

utterance to reply ask-wh-question-aboutagent is answer-wh-question-aboutagent, then the augmented version with a happiness emotion induced to the virtual human is answer-wh-question-aboutagent-posEmotion.

The second data collection was aimed to collect the interaction data about what type of behaviour people playing the *Smile Game* would exhibit and then implement those behaviours and interaction rules to the virtual human in *The Smile Game*. To collect the data, ten students were recruited to play the *Smile Game* with each other in pairs while being video and audio recorded. The interaction videos then were annotated using FACS coding system Ekman et al. (2002*a*). In a total of five sessions and a total of 18 games, sixty-two annotations were collected. The annotations were classified into two categories: Attack and Defend with 40 annotations belong to the Attack classification and 22 belong to the Defend group. Moreover, the interaction video also contributes a set of patterns of behaviour shown in this kind of the game. From there, the patterns were implemented as the virtual human's strategy to win the game by adopting a finite state's approach (see Figure 4.4).

Finally, additional interaction rules were also collected during the interaction between the players and the system in some *The Smile Game* studies to see what would the players talk about with a virtual human in a game setting. A total of additional 19 dialogue acts were extracted from a total of 871 annotations from the players and virtual human interaction in *The Smile Game* setting (see Table 5.5). The challenge to collect the data was how to simulate a natural interaction between players and virtual human. There were several challenges: First, using the WoZ technique potentially slowed down the flow of interaction due to the time taken by the Wizard to choose or type the best reply to the user. Second, the participant should interact naturally while also aware that they are being video recorded due to ethical reason. Third, some of the data were collected in the public space, where there were too many noise (i.e. sound and light), causing some of the data were unable to be processed due to the extensive noises. This might be an excellent dataset to extract voice using some approach similar to The Cocktail Party Problem (Haykin & Chen 2005).

8.1.3 Social Interactions with Game Companions

Games serve as a perfect tool to evaluate the affective loop systems (Yannakakis & Paiva 2014). In the game's virtual world, the players are continuously engaged and interacted with the game elements. During the interaction, the players are to be likely to elicit emotions as a response to the interaction with the virtual human in a game (i.e. NPC). There is no unified guideline to evaluate the believability of the virtual human as no unified set of qualities to define believability. Hence, the evaluation would take place in the user's opinions regarding the virtual human's behaviour, personality, the flow of the interaction, and the entire experience during the interaction (Bogdanovych et al. 2016, Lee & Heeter 2015). This thesis presents the evaluation of the user's experience when interacting with a virtual human in a game setting. Moreover, this thesis also demonstrates the exploration of virtual human and human social interaction in the wild. Chapter 3, 4, 6, 7 show the simulations and evaluation of social relationship models between a player and game companion based on the affective loop experience created between them. Moreover, in Chapter 3, 4, and 6, the players were evidently able to see and rate the relationship they have towards the virtual humans (i.e. game companions) immediately. There was a total of 115 participants who were interacted with the virtual human in a total of 215 interactions across the two scenarios. The first scenario was The Smile Game, where a human player who abides in the real world played with a virtual human who inhabits a virtual world. The second scenario was The Skyrim Game, where an avatar represented the human player, who played with a virtual human in the virtual world. The scenarios were designed to support the affective loop (Höök 2008) between human and virtual human in the laboratory, semi-public and public setting. In an affective loop system, the user should experience emotions constructed in the interaction. The emotions process also arise from social context during the interaction between them. Both scenarios provided evidence that there is an increase in player experience when playing the game with the additional influence of interactive emotion, personality and social relationships to the virtual human's behaviours. Furthermore, the players reported that they were more emotionally involved and attached with the characters and the game when the framework and models were activated in the game.

From the studies done in Chapter 6 and focus group discussion described in Chapter 7, the framework and computational models allow players to play not only using a conventional game controller but also use their facial expressions to control the choices, actions of the characters, and the story of the game. In Chapter 6, the players reported that the game with the framework and computational models activated in the game, where they were able to use their facial expressions as an additional game input to interact with the NPCs, resulting in the enhancement of their experiences when they were playing the game.

Lessons Learnt from The Smile Game

Chapter 5 provides a numerous number of affective loop experience data between human-virtual human in the wild. In the initial study of The Smile Game, the players and the virtual human interacted in a lab setting. The initial study was intended to be a pilot study to explore the potential of the models and framework as well as to explore the interaction loop between human and virtual human in a game setting. The players were able to perceive the personality set to the virtual human correctly. The players were also reported that they could see the social relationships with both virtual humans improved in the second encounter with them. This due to a simple strategy on remembering the player's face using facial expression recognition and called the player by their name. By doing so, it provides an illusion that the virtual human actually knew the player Hall (1966). From the technology, the virtual human received a relatively low score in their response time and naturalness in the conversation due to the ASR module but scored quite high in their naturalness when playing the game. The virtual human naturalness during the game was a statistically significant higher than in the conversation as in the game, the virtual human did not require to use the ASR module ($t = 7.32, p < 0.001, SD = 0.68$ and $t = 4.70, p < 0.001, SD = 0.96$ for Poppy and Spike respectively). This due to the microphone used in the pilot study was the built-in microphone from a webcam.

In the semi-public and public setting, the personality set to the virtual human was still consistently perceived correctly by the players. Most of the players perceived Poppy as a

nice and friendly person, while Spike as a rigid, rude, unpleasant character who was almost impossible to defeat in the game. Most of the players were happy to interact with Poppy but irritated by Spike when playing the game. This indicates that the personality models could be utilised as a mechanic for game levelling, where Poppy could be set as a first level "Boss" and Spike could be the last level "Boss" and in the middle of the game, the player could encounter the transition between Poppy's personality with Spike's personality. Up to the author's knowledge, there is no game exists yet that implements personality as a difficulty adjustment in the game. This thesis demonstrates that personality could be used as a variable to adjust the difficulty in the game (both The Skyrim Game and The Smile Game evaluation support this claim). Unfortunately, due to the result of the experimental setting of the semi-public and public space *Smile Game* studies and a relatively low turnout of the pilot studies, we could not evaluate the social relationship models. By nature, it is challenging to evaluate the social relationship variable in the semi-public and public space setting as there is a low probability that the same person would come and interact with the virtual human more than once. Unless the user/player is a regular in the place where we deploy the virtual human. Hence, a different setting is required to evaluate the social relationship models.

The players' arousal in the *semi-public* was statically significantly higher compare to the one in the *public* area , $Z = 7.3, p < 0.001$. While the players' valence was statically significantly higher in the *public* space, $Z = 5.73, p < 0.001$. This indicates, in this game, players were more attentive and expressive in a *semi-public* area. However, the players showed their enjoyment more when playing in a *public* space. As in the public area, the players were with their families or friends when playing the game. The players' valence and arousal were increasing at the beginning of the game and relatively steady in the middle throughout to the end of the game in both spaces, except for the valence in a *semi-public* space. In general, the peak of players' arousal and valence was seen in the almost end of the interaction before it went down approaching zero at the very end of the interaction. The interaction reached its climax when the players showed their best effort to make the virtual human smile as the music indicated that the game was approaching its end. Some of the limitations of the

system are: The system is only capable to process one player's facial expression and face recognition at a time; The system is currently not capable of dealing with occlusion (e.g. candies and ice cream) against the player's face; The system still not able to deal with noises, both sounds and lights; The social relationship module is only partially evidently evaluated in The Smile Game scenario due to the setting. Finally, there is an ethical issue should be raised in using this framework and models. To be able to recognise the user/player, the system is required to capture several pictures of the player's face and store it to the system. The pictures then are used as training data to be used in the next interaction. Hence, a consent form is needed to get permission to do the face capturing and storing process.

Lessons Learnt from The Skyrim Game

The second game scenario was The Skyrim game with a total of 60 interactions and a total play time of 18.16 hours and 2091.63 hours in real and game time respectively. The aims for the study was to investigate if imbuing affective loop during the interaction with the virtual human (i.e. NPC) in a game will increase the player's experience in the game. The study also demonstrates that personality and social relationship variable can be implemented as the game balancing and levelling system in the game. The quest of the game was to build a relationship with two characters in the Skyrim world and get a sword from them. The favourite strategy to build relationships with the NPCs was giving favourable items to them, with exploring Skyrim together as the second choice of strategy. Forty percent of players decided to go towards negative relation with Max as their game strategy to get The Sword. This indicates that different parameters of the variables in the models can be utilised as tools for game designers to design the game balancing and levelling. For example, High Neuroticism in the Personality model and low Social Relationship Intention character can be the ultimate challenge to the players to finish the quest.

The results clearly indicate that the game with the models introduced in the game provides more experiences to the players. With the models implemented, the players feel more engaged and emotionally immersed in the game and the characters when playing with both NPCs.

Players reported a significant increase in the engagement level with the game ($p < 0.01$). A significant difference was also found in the players' emotional attachment to the game and the NPCs with the models implemented ($p < 0.01$). Moreover, the models enriched the game story, increased the players' curiosity how the game would progress ($p < 0.01$). From the short interview, most of the players preferred the NPCs with models. The models provide an interesting flow to forge the relationships with NPCs.

Moreover, A Wilcoxon Signed-Ranks Test indicated that players' arousal was statically significantly higher with models implemented, $Z = 8.4, p < 0.001$ and $Z = 8.71, p < 0.001$, with Max and Stella respectively. In addition, players' valence was also significantly higher with models implemented, $Z = 3.1, p < 0.05$ and $Z = 7.75, p < 0.001$, with Max and Stella respectively. Negative valence also occurred in the interaction with model implemented Max. Players' arousal and valence were significantly higher when playing with model implemented Stella compare to Max, $Z = 3.75, p < 0.001$ and $Z = 6.47, p < 0.01$, respectively. On the contrary, there was no significant differences between players' arousal and valence when interacting with baseline Stella compare to Max, $Z = -1.674, p = 0.5$ and $Z = -1.6509, p = 0.1$, respectively.

8.1.4 Summary

This thesis provides empirical findings of a natural and believable interaction between a player and virtual human leading to create affective loop experience that is informed by the process of encoding and decoding of social signals exchanged between the player and game companion. (**RQ1**, shown in Chapter 3 - 7). The framework and computational models proposed and described in Chapter 3 allow the system to capture the user's (e.g. player's) emotions from their facial expression captured by a camera. The system was also capable of capturing the user's utterances from a microphone. The computational models of emotions, personality, and social relationship then process the user's emotions and utterances to update the system's internal states (e.g. emotions, social relationship perception towards the user) and propose the best response to be shown to the user. Chapter 4 to Chapter 6 show that the

proposed framework and computational models have been evaluated with several scenarios. The results indicate that in general, the majority of the participants were enjoyed to have the affective loop interaction with the system (Shown in Chapter 4 to Chapter 6). However, the system is limited to a very noisy environment, affecting the Automatic Speech Recognition Module. In Chapter 5, most of the participants broke their interaction with the system due to the system's ability to understand what the participants said to the system.

The repeated interaction between a player with the virtual human creating a natural and believable social relationship between them based on the affective loop experience (**RQ2**, demonstrated in Chapter 3, 4, 6, and 7). The computational models allow the system to mathematically compute the value of the system's (i.e. virtual human's) emotions and social relationship perception between the system and the user. Chapter 3 illustrates the simulation of the emotions and social relationship computational models with a sequence of events where a user was interacting with two virtual humans with a different personality. From the pilot study of *The Smile Game* (Chapter 4) and the Skyrim game (Chapter 6), players also reported that they could see what kind of relationship they have with the virtual human from the virtual human actions and attitudes towards them in the game. However, the limitation of the computational models proposed is that there are quite a number of variables to be adjusted to create the best value for each variable to a virtual human. Moreover, the user is required to interact with the system repeatedly to see the effect of the social relationship computational models to the interaction quality between them.

They also notified the changes in the relationship between them immediately (**RQ3**). In Chapter 4, the participants were able to tell the changes in their social relationship immediately in the second session of the game ($p < 0.001$ for both Poppy and Spike).However, the participants somehow did not see the relationships would change the quality of their interaction. In Chapter 6, the participants were also able to tell what relationship they have with a game companion in a game immediately from the questionnaire and their attitudes shown by their facial expressions, gestures, and verbal cues. In contrast with the Smile Game Scenario evaluated in Chapter 4, in this study, the changes in the social relationship had an effect to the participant's experience when playing the game, where there was a statically

significant improvement in the game experiences when the computational models activated in the game compared to the one with-out computational models activated.

Moreover, this thesis also presents on how to collect and craft the interaction rules for the virtual human in a specific game scenario (**RQ4**, illustrated in Chapter 4 and 5). Chapter 4 shows that the interaction rules were collected and annotated from the human-human interaction. The audio-visual recording data were annotated using the Facial Action Coding System (FACS) (Ekman et al. 2002*a*) for the participants' facial expressions as well as a transcription from the participants' utterances. Moreover, the interaction rules were also augmented using The SEMAINE Database Corpus (McKeown et al. 2012) annotation (with Dialogue Act annotation scheme (Austin 1975)). Chapter 5 also demonstrates the additional interaction rules collected from the participants' interaction with Poppy and Spike in the Pilot, WoZ, and Fully Automatic System.

From The Skyrim Game study, the players reported that they enjoyed more the game with the framework and models induced in the game allowing the player to use their facial expressions as an additional input to the game controller. The aspects enhanced in the player's experiences are the player's engagement and emotional attachment to the story, game and characters. The players also became more curious on how the game would progress (**RQ6**, shown in Chapter 6). However, not all game genre suitable to implement the framework, and models proposed and having the player's facial expressions as an additional input to the game controller. A game with rich-story, such as Role Playing Game or Visual Novel genre would be the best type of game to have this framework, and computational models implemented.

Finally, the series of the evaluation demonstrates the exploration of the human-virtual human interaction in the several settings (i.e. lab, semi-public and public) (**RQ5**). Chapter 4 and 6 shows the evaluation of the human-virtual human interaction in a lab scenario. Chapter 5 illustrates the evaluation of the interaction in semi-public and public settings. In a public setting, the participants showed statistically higher valence in the public space setting ($p < 0.001$) than in the lab or semi-public space setting. While, the participants' arousal was

statically significantly higher in the semi-public space setting, compared to the interaction in the public space setting ($p < 0.001$). Finally, in the public space, the system encountered a couple of problems related to the noise, lightning, and position of the system and the participants. However, the series of the evaluation provides a numerous number of human-virtual human interaction data to be explored (shown in Chapter 5), where there is not many research done in this area.

8.2 Possible Applications

With ERiSA framework and computational models, the VH designers will be able to design a VH that can perceive and exhibit emotions, has a unique personality and capable of building a simple relationship with the user. Several possible applications can be built by using the ERiSA framework and computational models. Following are the area of applications that the ERiSA framework models can be implemented:

8.2.1 Entertainment

The ERiSA models and framework provide a tool to design a game that provides more realistic interaction between game characters and players to a game designer. Personality model provides the uniqueness of each game character to enrich the game story and can be implemented as a game levelling system where characters with some personalities require more advanced strategy to deal with, compared to the others. Emotions model accommodate the emotional attachment to the characters leading to the increase of the game immersion and engagement. Social Relationship model supports the virtual human to forge a relationship with the players leading to the increase of the game engagement and experience. The designer can design a game with multiple endings depending on the social relationship status with each character.

Moreover, the ERiSA models and framework provides additional inputs for players to interact

with the game characters by using more natural modalities for interactions such as their facial expressions and spoken dialogue. Games with social and affect-driven characters such as Milo Project is now possible to be developed by implementing ERiSA models and framework. With the ERiSA models, players now can design not only the game character's outer layer of the outfit or appearances but also their personality. This is to say that the players will be able to create a character that represents themselves or someone they know in the virtual world. More examples of the potential of the ERiSA models and framework to design games are described in section 7.4.3. This would potentially shift the paradigm to design a game from a linear story scripting to a dynamic story scripting. Also, the designers would be able to explore new interactions form to enhance the player's experience.

The ERiSA models and framework can also be implemented to animate a physical embodiment by using a microcomputer such as Raspberry Pi to create an emotionally and socially aware toy with a unique personality. With the ERiSA models, the user can interact with the toy with limited dialogue and other non-verbal communications such as emotions from facial expressions, voice, and touch. Those patterns of interactions eventually will form a relationship between the user and the toy.

8.2.2 Business and Marketing

Some big companies have started to implement virtual human technology to help their operational activities primarily in the activity that involving customers for instances: virtual human for customer service, help centre, etc. However, the current technology is still far away to be called as a virtual human, as it is still a simple chat-bot with a knowledge of only about a specific topic (e.g. information about the company, information about the products and FAQs).

With the ERiSA framework and models, the designers now can induce emotions, personality and the ability to build a social relationship with the customers, to the virtual human. The designer also can introduce models of empathy to the virtual human, where it can be a key to build a relationship with the customers. If the relationship is successful, the company

will be able to easily market their product to the customers by persuading them to buy a new product or recommending them a new product based on the customers' preferences. Moreover, the virtual human can also be used as a good tool to collect customers' data more naturally compare to filling in the survey. The virtual human can interact with the customers naturally while collecting their data such as what is their favourite food, colours, where do the customers usually buy their milk, etc.

8.2.3 Medical

The models and framework can be used as a tool to build a virtual human for training purposes in the medical area. Virtual human implementation to the medical area, in general, has a great advantage. It provides a real-life simulation to the doctors or nurses for training purposes without potentially risking the patients' lives. Moreover, previous studies show that patients were more comfortable to have a conversation with a virtual therapist compare to the real one as a virtual therapist would not judge them nor having prejudice toward them.

The models provide the opportunity for doctors or nurses to train with a virtual human with different personalities. The traits can be replicated from real patients. Moreover, with the models and framework, the doctor and nurses can train on how to build social relationships with the patients through a virtual human. The designers can build a specific scenario based on a real medical case. By following the methods describes in this thesis on how to collect interaction rules based on real interaction video, the designers should be able to replicate the method and implement it to the medical scenarios. However, this should be done with the supervision of the experts in the field.

8.2.4 Education and Training

In education, a virtual human is generally used as training and assessment tools because they have the unique ability to engage with students. The models provide the capability to design a realistic but limited simulation on how people should behave based on a specific event designed to a specific scenario. The framework also provides the capability to read the player's facial expressions and mapped them to the games. For examples: a simulation to teach children how to behave appropriately in public. Children will get points if they can display appropriate behaviours, evaluated by their facial expressions, emotions from voice and choices, given a specific scenario. Another example is a serious game for autism players. With the supervision of the experts, the game will be a potential tool to help the player to understand and display the appropriate emotions. The player's scores or points, for example, will be reduced if they can not identify the correct emotions displayed or are not able to reproduce the same emotions acted by the game characters.

For a teacher, the models and framework can be used to create virtual students/pupils to interact with. The virtual students can be designed to have a unique personality. Hence the teacher can learn how to interact with each student and build a relationship with them leading to trust. Theory of motivation and goal can be induced to the models to create a more realistic virtual student, so each virtual student will have a different goal/motivation to study.

8.3 Future Work

There are some problems remain to be solved and need to be addressed in future work. The author aware that the system is still not ready to run fully automatically in a *public* space, as the environment, can be very noisy, affecting the sensors (e.g. microphone for automatic speech and emotions recognition and camera for non-verbal behaviours analysis). As the framework is a collection of integrated and modular tools, it is indeed quite tricky to get the system to run smoothly. A flaw in one tool can cause to the decrease of the virtual

human believability. The most problematic problem is the ASR module. However, there are some solutions to this problem. Either implement an excellent commercial ASR API that provides better models for speech such as Nuance or Google Speech API or builds and train the models in a specific scenario. In addition, a trained model of noise in the background can be implemented to filter the noise.

This thesis only demonstrates a game system that interprets emotions from the player's facial cues. However, in a real-world scenario, human interpret emotions through multi-modal cues. Though the framework allows the designers to interpret emotions by using multi-modal cues, there is still an ongoing question on how to fuse the information from multiple cues. For example: which cues hold the essential information for the emotions and if the information from each cue contradict each other, which information should the system use? Also, in which level should the fusion take part (low, medium, high). The level fusion would need to be studied (Valstar 2008). It could be different for each scenario depending on the purpose of the application.

The interaction rules in this thesis are still manually hand-crafted. This would require an enormous number of time and labour to create the rules. With machine learning, statistical models can be designed to learn the interaction rules automatically. Several scenarios can be designed to collect the data. For example, the scenarios demonstrated in this thesis: human-human-computer interactions (or human-human interaction with machine-mediated) and Wizard of Oz with human and virtual human interaction with the real system or human. Moreover, the interaction rules crafted in this thesis would only work best with specific scenarios described in this thesis as there are insufficient data to generalise the behaviours to all possible scenarios. Additional rules generation might be required for each scenario. However, the interaction rules described in this thesis might be re-used in a similar scenario (e.g. RPG games).

The scenarios in this thesis only implemented to a portion of a game (i.e. only several levels or quests). The models and framework can be evaluated in a full-scale game. The models can be implemented as part of the game design to enrich the game story or can be used to

create a serious game. This thesis demonstrates game scenarios where the player is in the real world while the virtual human is in the virtual world and both player and virtual human are inhibiting the virtual world. It would be interesting if we can bring the NPCs to the real world and playing with the player in the real world. This would be possible as the model and framework also allow the implementation into a physical embodiment by using affordable tools like raspberry pi to a sophisticated robot such as NAO[1] that can be controlled through BML (Le & Pelachaud 2012).

[1]http://www.robotlab.com/

Appendix A

Attacks and Defences

Table A.1: List of Additional Jokes

UTTERANCES (level)
Hello, love (1)
For the love of God (1)
In the beginning, God created the heavens and the earth (1)
The voices in my head may not be real, but they have some good ideas! (2)
Never, under any circumstances, take a sleeping pill and a laxative on the same night (2)
Don't you have something better to do? (3)
I have two sons; both are boys (3)
I have two daughters; both are girls! (3)
They who laughs last thinks slowest(3)
You know you are fat when you hug a child, and it gets lost. (3)



Figure A.1: Attack Behaviours Lv. 1



Figure A.2: Attack Behaviours Lv. 2



Figure A.3: Attack Behaviours Lv. 3



Figure A.4: Defend Behaviours Lv. 1



Figure A.5: Defend Behaviours Lv. 2



Figure A.6: Defend Behaviours Lv. 3

Appendix B

Evaluation Forms

SMILE GAME EVALUATION FORM

DATE		AGE		ETHNICITY	
SESSION		GENDER			

POPPY

1 On a scale of 1-5 how would you rate your relationship with Poppy?

No Relationship	Not Close	Quite Close	Close	Very Close

2 On a scale 1-5 how would you view the following characteristics of Poppy?

a Extraversion (Energy, positive emotions, talkativeness)

Very Low	Low	Medium	High	Very High

b Neuroticism (Sensitive/nervous, negative emotions, anger, anxiety)

Very Low	Low	Medium	High	Very High

3.a How natural was your interaction with Poppy during the conversation?

Very Unnatural	Unnatural	Average	Natural	Very Natural

b On a scale 1-5 how would you rate Poppy's response time during the conversation?

Very Slow	Slow	Average	Fast	Very Fast

4 How natural was your interaction with Poppy during the game?

Very Unnatural	Unnatural	Average	Natural	Very Natural

SPIKE

5 On a scale of 1-5 how would you rate your relationship with Spike?

No Relationship	Not Close	Quite Close	Close	Very Close

6 On a scale 1-5 how would you view the following characteristics of Spike?

a Extraversion (Energy, positive emotions, talkativeness)

Very Low	Low	Medium	High	Very High

b Neuroticism (Sensitive/nervous, negative emotions, anger, anxiety)

Very Low	Low	Medium	High	Very High

7.a How natural was your interaction with Spike during the conversation?

Very Unnatural	Unnatural	Average	Natural	Very Natural

b On a scale 1-5 how would you rate Spike's response time during the conversation?

Very Slow	Slow	Average	Fast	Very Fast

8 How natural was your interaction with Spike during the game?

Very Unnatural	Unnatural	Average	Natural	Very Natural

Figure B.1: Evaluation Form for The Smile Game Pilot Study

The University of Nottingham

THE SKYRIM GAME PRE-QUESTIONNAIRE

Name _____ Ethnicity _____

Age _____ Date _____

Gender _____ Played Skyrim before **YES** **NO**

Instructions

In the table below, for each statement 1-50 mark how much you agree with on the scale 1-5, where 1=disagree, 2=slightly disagree, 3=neutral, 4=slightly agree and 5=agree, in the box to the left of it.

Rating	I....	Rating	I....
	1. Am the life of the party.		26. Have little to say.
	2. Feel little concern for others.		27. Have a soft heart.
	3. Am always prepared.		28. Often forget to put things back in their proper place.
	4. Get stressed out easily.		29. Get upset easily
	5. Have a rich vocabulary		30. Do not have a good imagination
	6. Don't talk a lot		31. Talk to a lot of different people at parties.
	7. Am interested in people		32. Am not really interested in others
	8. Leave my belongings around		33. Like order.
	9. Am relaxed most of the time.		34. Change my mood a lot.
	10. Have difficulty understanding abstract ideas		35. Am quick to understand things
	11. Feel comfortable around people.		36. Don't like to draw attention to myself
	12. Insult people.		37. Take time out for others.
	13. Pay attention to details.		38. Shirk my duties
	14. Worry about things.		39. Have frequent mood swings.
	15. Have a vivid imagination		40. Use difficult words.
	16. Keep in the background.		41. Don't mind being the center of attention.
	17. Sympathize with others' feelings.		42. Feel others' emotions.
	18. Make a mess of things.		43. Follow a schedule
	19. Seldom feel blue.		44. Get irritated easily
	20. Am not interested in abstract ideas		45. Spend time reflecting on things.
	21. Start conversations.		46. Am quiet around strangers.
	22. Am not interested in other people's problems.		47. Make people feel at ease.
	23. Get chores done right away.		48. Am exacting in my work
	24. Am easily disturbed.		49. Often feel blue.
	25. Have excellent ideas		50. Am full of ideas

Figure B.2: Pre Questionnaire Form for The Skyrim Study

THE SKYRIM GAME QUESTIONNAIRE - STELLA

The University of Nottingham

Name _____ Session _____

Date _____

Do you see Stella as someone who ...		Strongly Disagree	A little Disagree	Neither Agree nor Disagree	A little Agree	Strongly Agree
1	positive thinking					
2	negative thinking					
3	generates a lot of enthusiasm					
4	tends to be quiet					
5	is outgoing, sociable					
6	is depressed, blue					
7	is relaxed, handles stress well					
8	can be tense					
9	worries a lot					
10	is emotionally stable, not easily upset					
11	can be moody					
12	is sometimes rude to others					

		Strongly Disagree	A little Disagree	Neither Agree nor Disagree	A little Agree	Strongly Agree
13	Stella is someone that you can count on to listen when you are very angry at someone else					
14	I really like Stella					
15	I am familiar with Stella					
16	I am close with Stella					

I fell .... when I interact with Stella		Very slightly or not at all	a little	Moderately	quite a bit	extreamely
17	Happy					
18	Angry					
19	Sad					
20	Afraid					
21	Disgust					
22	Surprise					

		Strongly Disagree	A little Disagree	Neither Agree nor Disagree	A little Agree	Strongly Agree
23	Playing seems automatic					
24	I lose track of where I am					
25	I play without thinking about how to play					
26	I feel like I just can't stop playing					
27	I really get into the game					
28	If someone talks to me, I don't hear them					
29	I felt that I really empathised/felt for with the game					
30	I did not feel any emotional attachment to the game					
31	I was interested in seeing how the game's events would progress					
32	I sometimes found myself to become so involved with the game that I wanted to speak to the game directly					

Figure B.3: Pre and Post Stella's Questionnaire Form for The Skyrim Study

THE SKYRIM GAME QUESTIONNAIRE - MAX

The University of Nottingham

Name _____ Session _____

Date _____

	Do you see Max as someone who ...	Strongly Disagree	A little Disagree	Neither Agree nor Disagree	A little Agree	Strongly Agree
1	positive thinking					
2	negative thinking					
3	generates a lot of enthusiasm					
4	tends to be quiet					
5	is outgoing, sociable					
6	is depressed, blue					
7	is relaxed, handles stress well					
8	can be tense					
9	worries a lot					
10	is emotionally stable, not easily upset					
11	can be moody					
12	is sometimes rude to others					

		Strongly Disagree	A little Disagree	Neither Agree nor Disagree	A little Agree	Strongly Agree
13	Max is someone that you can count on to listen when you are very angry at someone else					
14	I really like Max					
15	I am familiar with Max					
16	I am close with Max					

	I fell .... when I interact with Max	Very slightly or not at all	a little	Moderately	quite a bit	extreamely
17	Happy					
18	Angry					
19	Sad					
20	Afraid					
21	Disgust					
22	Surprise					

		Strongly Disagree	A little Disagree	Neither Agree nor Disagree	A little Agree	Strongly Agree
23	Playing seems automatic					
24	I lose track of where I am					
25	I play without thinking about how to play					
26	I feel like I just can't stop playing					
27	I really get into the game					
28	If someone talks to me, I don't hear them					
29	I felt that I really empathised/felt for with the game					
30	I did not feel any emotional attachment to the game					
31	I was interested in seeing how the game's events would progress					
32	I sometimes found myself to become so involved with the game that I wanted to speak to the game directly					

Figure B.4: Pre and Post Max's Questionnaire Form for The Skyrim Study

The University of Nottingham

WORKSHOP QUESTIONNAIRE

Name _____ Ethnicity _____

Age _____ Date _____

Gender _____ How long have you been in Game Industry ? _____

Instructions

In the table below, for each statement 1-50 mark how much you agree with on the scale 1-5, where 1=disagree, 2=slightly disagree, 3=neutral, 4=slightly agree and 5=agree, in the box to the left of it.

Rating	I....	Rating	I....
	1. Am the life of the party.		26. Have little to say.
	2. Feel little concern for others.		27. Have a soft heart.
	3. Am always prepared.		28. Often forget to put things back in their proper place.
	4. Get stressed out easily.		29. Get upset easily
	5. Have a rich vocabulary		30. Do not have a good imagination
	6. Don't talk a lot		31. Talk to a lot of different people at parties.
	7. Am interested in people		32. Am not really interested in others
	8. Leave my belongings around		33. Like order.
	9. Am relaxed most of the time.		34. Change my mood a lot.
	10. Have difficulty understanding abstract ideas		35. Am quick to understand things
	11. Feel comfortable around people.		36. Don't like to draw attention to myself
	12. Insult people.		37. Take time out for others.
	13. Pay attention to details.		38. Shirk my duties
	14. Worry about things.		39. Have frequent mood swings.
	15. Have a vivid imagination		40. Use difficult words.
	16. Keep in the background.		41. Don't mind being the center of attention.
	17. Sympathize with others' feelings.		42. Feel others' emotions.
	18. Make a mess of things.		43. Follow a schedule
	19. Seldom feel blue.		44. Get irritated easily
	20. Am not interested in abstract ideas		45. Spend time reflecting on things.
	21. Start conversations.		46. Am quiet around strangers.
	22. Am not interested in other people's problems.		47. Make people feel at ease.
	23. Get chores done right away.		48. Am exacting in my work
	24. Am easily disturbed.		49. Often feel blue.
	25. Have excellent ideas		50. Am full of ideas

Figure B.5: Pre Questionnaire Form for Design Workshop

Appendix C

XML Format for The Framework

Listing C.1: Rules for Intentions and States

```
<?xml version="1.0"?>
<rules>
 <rule id="R1">
  <words>
   <word value="hello"/>
  </words>
  <states>
   <state name="$playerstates.intention" value="greeting"/>
   <state name="$dialoguestates.topic" value="greeting"/>
  </states>
 </rule>
 <rule id="R2">
  <words>
   <word value="hello *"/>
  </words>
  <states>
   <state name="$playerstates.intention" value="greeting"/>
   <state name="$dialoguestates.topic" value="greeting"/>
  </states>
 </rule>
</rules>
```

Listing C.2: Game Companion's Data

```
<?xml version="1.0"?>
<actors>
  <agent id="A1">
    <firstname>Poppy</firstname>
    <lastname>Williams</lastname>
    <datebirth>4</datebirth>
    <monthbirth>5</monthbirth>
    <yearbirth>1852</yearbirth>
    <gender>female</gender>
    <openness>0.0</openness>
    <conscientiousness>0.0</conscientiousness>
    <extraversion>0.5</extraversion>
    <agreeableness>0.0</agreeableness>
    <neuroticism>0.0</neuroticism>
    <pWeight>0.02</pWeight>
  </agent>
</actors>
```

Listing C.3: Agent's Utterances

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<utterances>
  <agent name="Poppy">
    <category name="greeting">
      <phrase id="1" value="Hello $NAME, how are you?"/>
      <phrase id="2" value="Hi $NAME, How was your day?"/>
      <phrase id="3" value="Hello $NAME, welcome to SMILE GAME"/>
      <phrase id="4" value="Hi $NAME, How's everything going?"/>
      <phrase id="5" value="Hello $NAME, Good to see you."/>
    </category>
    <category name="respondpos">
      <phrase id="1" value="Amazing!"/>
      <phrase id="2" value="That's wonderful!"/>
      <phrase id="3" value="That's great!"/>
      <phrase id="4" value="I'm happy to hear that!"/>
```

```
16        <phrase id="5" value="Fantastic, I love to hear it!"/>
17      </category>
18      <category name="ask-wh-question-aboutusername">
19        <phrase id="1" value="Hi,you look new for me.What's your name?"/>
20        <phrase id="2" value="Hi,What was your name again?"/>
21        <phrase id="3" value="Hi,What's your name?"/>
22        <phrase id="4" value="How may I address you?"/>
23        <phrase id="5" value="Hello there,What's your name?"/>
24      </category>
25  </utterances>
```

Listing C.4: Agent's Background Story

```
1  <?xml version="1.0" encoding="UTF-8" standalone="no" ?>
2  <bgStories>
3
4    <agent value="Poppy">
5      <bgStory id="BG1" subject="self" location="Salisbury" topic="
           aboutAgent" day="8" month="3" year="1987" affinity="1.00">
6        <story id="1" value="$SUBJECT was born at small city called $
             LOCATION" />
7        <story id="2" value="$SUBJECT was born on $DATE" />
8      </bgStory>
9
10     <bgStory id="BG2" subject="self" location="Salisbury" topic="aboutGame
           " day="22" month="11" year="1995" affinity="0.90">
11       <story id="1" value="My mom introduced $SUBJECT to The Smile Game on
             $DATE" />
12       <story id="2" value="$SUBJECT learned to play The Smile Game when I
             was a kid. It was my mom who introduced me to the game." />
13     </bgStory>
14   </agent>
15
16   <agent value="Spike">
17     <bgStory id="BG1" subject="self" location="Sheffield" topic="
           aboutAgent" day="3" month="8" year="1988" affinity="0.80">
```

```
18        <story id="1" value="$SUBJECT was born at $LOCATION" />
19        <story id="2" value="$SUBJECT was born on $DATE" />
20      </bgStory>
21
22      <bgStory id="BG2" subject="self" location="London" topic="aboutGame"
            day="16" month="8" year="1998" affinity="0.70">
23        <story id="1" value="My friend Tom introduced $SUBJECT to The Smile
              Game on $DATE" />
24        <story id="2" value="$SUBJECT learned to play The Smile Game when I
              was in primary school. My classmate Tom was the one who
              introduced me." />
25      </bgStory>
26    </agent>
27
28 </bgStories>
```

Listing C.5: Social Relationships Data

```
1 <?xml version="1.0" encoding="UTF-8" standalone="no"?>
2 <SocialRelationships>
3   <relation actorA="A1" actorB="P1" id="R1" value="0.05">
4     <affinity>0.5</affinity>
5     <familiarity>1.0</familiarity>
6     <dateFirstMet>10</dateFirstMet>
7     <monthFirstMet>7</monthFirstMet>
8     <yearFirstMet>2014</yearFirstMet>
9     <hourLastMet>11</hourLastMet>
10    <minuteLastMet>2</minuteLastMet>
11    <secondLastMet>35</secondLastMet>
12    <dateLastMet>18</dateLastMet>
13    <monthLastMet>1</monthLastMet>
14    <yearLastMet>2016</yearLastMet>
15    <maxFamiliarity>1000.0</maxFamiliarity>
16    <decayRate>0.12</decayRate>
```

```
17      <halfLife>0.12</halfLife>
18      <intentionAffinity>0.02</intentionAffinity>
19      <interactionNumber>10.0</interactionNumber>
20    </relation>
21 </SocialRelationships>
```

Listing C.6: Known Faces Database

```
1 <?xml version="1.0" encoding="UTF-8" standalone="no" ?>
2 <players>
3   <player id="1">
4     <name first="Andry" last="Chowanda"/>
5     <birth date="8" month="3" year="1987"/>
6     <face id="1" path="/Models/CSV/FaceDB/CHO0.JPG"/>
7     <face id="2" path="/Models/CSV/FaceDB/CHO1.JPG"/>
8     <face id="3" path="/Models/CSV/FaceDB/CHO2.JPG"/>
9     <face id="4" path="/Models/CSV/FaceDB/CHO3.JPG"/>
10    <face id="5" path="/Models/CSV/FaceDB/CHO4.JPG"/>
11    <face id="6" path="/Models/CSV/FaceDB/CHO5.JPG"/>
12    <face id="7" path="/Models/CSV/FaceDB/CHO6.JPG"/>
13    <face id="8" path="/Models/CSV/FaceDB/CHO7.JPG"/>
14    <face id="9" path="/Models/CSV/FaceDB/CHO8.JPG"/>
15    <face id="10" path="/Models/CSV/FaceDB/CHO9.JPG"/>
16  </player>
17 </players>
```

[This page intentionally left blank]

Bibliography

AB, V. T. (2016), *MPEG-4 Face and Body Animation*, Visage Technology AB, Tegskiftesgatan 291, SE-58334 Linkping, Sweden.

Alden, D. L., Hoyer, W. D. & Lee, C. (1993), 'Identifying global and culture-specific dimensions of humor in advertising: A multinational analysis', *The Journal of Marketing* pp. 64–75.

Almaev, T., Martinez, B. & Valstar, M. (2015), Learning to transfer: transferring latent task structures and its application to person-specific facial action unit detection, *in* 'Proceedings of the IEEE International Conference on Computer Vision', pp. 3774–3782.

Almaev, T. R., Yüce, A., Ghitulescu, A. & Valstar, M. F. (2013), Distribution-based iterative pairwise classification of emotions in the wild using lgbp-top, *in* 'Proceedings of the 15th ACM on International conference on multimodal interaction', ACM, pp. 535–542.

Almaev, T. & Valstar, M. (2013), Local gabor binary patterns from three orthogonal planes for automatic facial expression recognition, *in* 'Affective Computing and Intelligent Interaction (ACII)', pp. 356–361.

Altman, I. & Taylor, D. A. (1973), *Social penetration: The development of interpersonal relationships.*, Holt, Rinehart & Winston.

Anderson, J. R. (1996), 'Act: A simple theory of complex cognition.', *American Psychologist* **51**(4), 355.

Andreatta, P., Klotz, J. J., Madsen, J. M., Hurst, C. G. & Talbot, T. B. (2015), 'Outcomes

From Two Forms of Training for First-Responder Competency in Cholinergic Crisis Management', *Military Medicine* **180**(4), 468–474.

Andrighetto, G., Governatori, G., Noriega, P. & van der Torre, L. W. N., eds (2013), *Normative Multi-Agent Systems*, Vol. 4 of *Dagstuhl Follow-Ups*, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik.

Arnold, M. B. (1960), 'Emotion and personality.'.

Austin, J. L. (1975), *How to do things with words*, Oxford university press.

Aylett, R., Paiva, A., Dias, J., Hall, L. & Woods, S. (2009), Affective agents for education against bullying, *in* J. Tao & T. Tan, eds, 'Affective Information Processing', Springer London, pp. 75–90.

Barbalet, J. M. (1996), 'Social emotions: confidence, trust and loyalty', *International Journal of Sociology and Social Policy* **16**(9/10), 75–96.

Bates, J. (1994), 'The role of emotion in believable agents', *Commun. ACM* **37**(7), 122–125.

Bevacqua, E., Prepin, K., Niewiadomski, R., de Sevin, E. & Pelachaud, C. (2010), GRETA : Towards an Interactive Conversational Virtual Companion, *in* 'Artificial Companions in Society: perspectives on the Present and Future', pp. 1–17.

Bevacqua, E., Sevin, E., Hyniewska, S. & Pelachaud, C. (2012), 'A listener model: introducing personality traits', *Journal on Multimodal User Interfaces* **6**(1-2), 27–38.

Bickmore, T. & Cassell, J. (2005), Social dialongue with embodied conversational agents, *in* 'Advances in natural multimodal dialogue systems', Springer, pp. 23–54.

Bickmore, T., Pfeifer, L. & Schulman, D. (2011), Relational agents improve engagement and learning in science museum visitors, *in* 'Proceedings of the 10th International Conference on Intelligent Virtual Agents', IVA'11, Springer-Verlag, Berlin, Heidelberg, pp. 55–67.
URL: *http://dl.acm.org/citation.cfm?id=2041666.2041675*

Bickmore, T., Vardoulakis, L., Jack, B. & Paasche-Orlow, M. (2013), Automated promotion of technology acceptance by clinicians using relational agents, *in* 'Intelligent Virtual Agents', Vol. 8108, Springer Berlin Heidelberg, pp. 68–78.

Bickmore, T. W., Pfeifer, L., Schulman, D., Perera, S., Senanayake, C. & Nazmi, I. (2008), Public displays of affect: deploying relational agents in public spaces, *in* 'CHI'08 Extended Abstracts on Human Factors in Computing Systems', ACM, pp. 3297–3302.

Bickmore, T. W., Rosalind & Picard, W. (2005), 'Establishing and maintaining long-term human-computer relationships', *ACM Transactions on Computer Human Interaction* **12**, 293–327.

Bogdanovych, A., Trescak, T. & Simoff, S. (2016), 'What makes virtual agents believable?', *Connection Science* **28**(1), 83–108.

Bothell, D. (n.d.), *ACT-R 6.0 Reference Manual*.

Bourne, L. E., Dominowski, R. L. & Loftus, E. F. (1979), *Cognitive Pcrocesses, Prentice-Hall series in experimental psychology*, Prentice-Hall.

Brockmyer, J. H., Fox, C. M., Curtiss, K. A., McBroom, E., Burkhart, K. M. & Pidruzny, J. N. (2009), 'The development of the game engagement questionnaire: A measure of engagement in video game-playing', *Journal of Experimental Social Psychology* **45**(4), 624–634.

Broekens, J., Hudlicka, E. & Bidarra, R. (2016), *Emotional Appraisal Engines for Games*, Springer International Publishing, pp. 215–232.

Brugman, H., Russel, A. & Nijmegen, X. (2004), Annotating multi-media/multi-modal resources with elan., *in* 'LREC'.

Buckland, M. (2005), *Programming game AI by example*, Jones & Bartlett Learning.

Burnett, S., Bird, G., Moll, J., Frith, C. & Blakemore, S.-J. (2009), 'Development during adolescence of the neural processing of social emotion', *Journal of cognitive neuroscience* **21**(9), 1736–1750.

Cacioppo, J. T. & Berntson, G. G. (1994), 'Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates.', *Psychological bulletin* **115**(3), 401.

Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsson, H. & Yan, H. (2000), Embodied conversational agents, MIT Press, Cambridge, MA, USA, chapter Human Conversation As a System Framework: Designing Embodied Conversational Agents, pp. 29–63.

Cavazza, M. (2000), 'Al in computer games: Survey and perspectives', *Virtual Reality* **5**(4), 223–235.

Cohn, J. F. (2007), Foundations of human computing: Facial expression and emotion, *in* 'Proceedings of the ICMI 2006 and IJCAI 2007 International Conference on Artifical Intelligence for Human Computing', ICMI'06/IJCAI'07, Springer-Verlag, Berlin, Heidelberg, pp. 1–16.
URL: *http://dl.acm.org/citation.cfm?id=1894917.1894919*

Dahlbäck, N., Jönsson, A. & Ahrenberg, L. (1993), Wizard of oz studies: Why and how, *in* 'Proceedings of the 1st International Conference on Intelligent User Interfaces', IUI '93, ACM, New York, NY, USA, pp. 193–200.

Damian, I., Endra, B., Huber, P., Bee, N. & Andr, E. (2011), Individualized agent interactions., *in* J. M. Allbeck & P. Faloutsos, eds, 'MIG', Vol. 7060 of *Lecture Notes in Computer Science*, Springer, pp. 15–26.

Dastani, M., Meyer, J.-J. C. & Grossi, D. (2013), 'A logic for normative multi-agent programs', *J. Log. Comput.* **23**(2), 335–354.

de Rivera, J. & Grinkis, C. (1986), 'Emotions as social relationships', *Motivation and Emotion* **10**(4), 351–369.

Descartes, R. (1988), *Meditations on First Philosophy*, NuVision Publications, LLC.

Descartes, R., Miller, V. & Miller, R. (1984), *Principles of Philosophy:*, A Pallas paperback,

Springer Netherlands.

URL: *http://books.google.co.uk/books?id=8vt0uWWVzEYC*

DeVault, D., Mell, J. & Gratch, J. (2015), Toward Natural Turn-Taking in a Virtual Human Negotiation Agent, *in* 'AAAI Spring Symposium on Turn-taking and Coordination in Human-Machine Interaction', AAAI Press, Palo Alto, California, pp. 2–9.

Dias, J., Mascarenhas, S. & Paiva, A. (2011), Fatima modular: Towards an agent architecture with a generic appraisal framework, *in* 'International Workshop on Standards for Emotion Modeling'.

Dweck, C. S. & Leggett, E. L. (1988), 'A social-cognitive approach to motivation and personality.', *Psychological review* **95**(2), 256.

Ekman, P. (1999*a*), *Basic emotions*, John Wiley & Sons Ltd, New York.

Ekman, P. (1999*b*), *Facial expressions*, John Wiley & Sons Ltd, New York.

Ekman, P. (2007), *Emotions revealed: Recognizing faces and feelings to improve communication and emotional life*, Macmillan.

Ekman, P. & Friesen, W. (1969), 'The repertoire of nonverbal behavior: Categories, origins, usage and coding', *Semiotica* **1**(1), 49–98.

Ekman, P., Friesen, W. & Hager, J. (2002*a*), *Facial Action Coding System (FACS): Manual*, A Human Face, Salt Lake City (USA).

Ekman, P., Friesen, W. V. & Hager, J. C. (2002*b*), *Facial Action Coding System: The Manual on CD ROM*, Research Nexus division of Network Information Research Corporation.

Elliott, C. D. (1992), The Affective Reasoner: A Process Model of Emotions in a Multi-agent System, PhD thesis, Evanston, IL, USA. UMI Order No. GAX92-29901.

Eyben, F., Wöllmer, M. & Schuller, B. (2010), Opensmile: the münich versatile and fast open-source audio feature extractor, *in* 'Proceedings of the international conference on Multimedia', ACM, pp. 1459–1462.

Eysenck, H. J. & Eysenck, S. (1965), 'The eysenck personality inventory'.

Eysenck, H. J. & Eysenck, S. B. (1976), *Eysenck personality questionnaire*, Educational and industrial testing service.

Fernández-Carbajales, V., Martínez, J. M. & Morán, F. (2006), High-level description tools for humanoids, *in* 'Proceedings of the 2006 International Conference on Multimedia Content Representation, Classification and Security', MRCS'06, Springer-Verlag, Berlin, Heidelberg, pp. 387–394.

Fetter, W. (1982), 'A progression of human figures simulated by computer graphics', *IEEE Computer Graphics and Applications* **2**(9), 9–13.

Fine, G. A. (1983), *Sociological Approaches to the Study of Humor*, Springer New York, New York, NY.

Foa, U. G. (1961), 'Convergences in the analysis of the structure of interpersonal behavior.', *Psychological Review* **68**(5), 341.

Frijda, N. H. (1988), 'The laws of emotion.', *American psychologist* **43**(5), 349.

Funder, D. C. (2001), 'Personality', *Annual Review of Psychology* **52**(1), 197–221.

Geijtenbeek, T. (2013), Animating Virtual Characters using Physics-Based Simulation, Phd thesis, Utrecht University.

Green, P. & Wei-Haas, L. (1985*a*), *The Wizard of Oz: a tool for rapid development of user interfaces.*

Green, P. & Wei-Haas, L. (1985*b*), *The Wizard of Oz: a tool for rapid development of user interfaces.*

Guise, V., Chambers, M. & Välimäki, M. (2012), 'What can virtual patient simulation offer mental health nursing education?', *Journal of psychiatric and mental health nursing* **19**(5), 410–418.

Gunes, H., Schuller, B., Pantic, M. & Cowie, R. (2011), Emotion representation, analysis and synthesis in continuous space: A survey, *in* 'Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on', pp. 827–834.

H-Anim (2016), 'H-anim'.
URL: *http://h-anim.org//*

Haag, A., Goronzy, S., Schaich, P. & Williams, J. (2004), Emotion recognition using bio-sensors: First steps towards an automatic system, *in* E. Andre, L. Dybkjr, W. Minker & P. Heisterkamp, eds, 'Affective Dialogue Systems', Vol. 3068 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 36–48.

Hall, E. (1966), *The hidden dimension*, A Doubleday anchor book, Doubleday.
URL: *http://books.google.co.uk/books?id=zGYPwLj2dCoC*

Hart, J., Gratch, J. & Marsella, S. C. (2013), How virtual reality training can win friends and influence people, *in* C. Best, G. Galanis, J. Kerry & R. Sottilare, eds, 'Fundamental Issues in Defense Training and Simulation', Human Factors in Defense, Ashgate.

Hartholt, A., Traum, D., Marsella, S. C., Shapiro, A., Stratou, G., Leuski, A., Morency, L. & Gratch, J. (2013), All together now: Introducing the virtual human toolkit, *in* 'International Conference on Intelligent Virtual Humans'.

Haykin, S. & Chen, Z. (2005), 'The cocktail party problem', *Neural computation* **17**(9), 1875–1902.

Hays, R. B. (1989), 'The day-to-day functioning of close versus casual friendships', *Journal of Social and Personal Relationships* **6**(1), 21–37.

Hennig-Thurau, T., Gwinner, K. P. & Gremler, D. D. (2002), 'Understanding relationship marketing outcomes: an integration of relational benefits and relationship quality', *Journal of service research* **4**(3), 230–247.

Heylen, D., Kopp, S., Marsella, S. C., Pelachaud, C. & Vilhjálmsson, H. (2008), The next step

towards a function markup language, *in* 'Proceedings of the 8th International Conference on Intelligent Virtual Agents', IVA '08, Springer-Verlag, Berlin, Heidelberg, pp. 270–280.

Heylen, D., Kopp, S., Marsella, S., Pelachaud, C. & Vilhjlmsson, H. (2008), The next step towards a function markup language, *in* 'Intelligent Virtual Agents', Vol. 5208, Springer Berlin Heidelberg, pp. 270–280.

Hickey, L. P. (2016), 'The brain in a vat argument'.
URL: *http://www.iep.utm.edu/brainvat/*

Hilgard, E. (1953), *Introduction to psychology*, Harcourt, Brace.

Ho, W. C. & Dautenhahn, K. (2008), Towards a narrative mind: The creation of coherent life stories for believable virtual agents, *in* 'International Workshop on Intelligent Virtual Agents', Springer, pp. 59–72.

Hoekstra, C. (2006), 'Adaptive artificially intelligent agents in video games: A survey', *UNIAI-06* .

Höök, K. (2008), Affective loop experiences-what are they?, *in* 'PERSUASIVE', Springer, pp. 1–12.

Hudlicka, E. (2004), Two sides of appraisal: Implementing appraisal and its consequences within a cognitive architecture, *in* 'Proceedings of the AAAI Spring Symposium 2004, Architectures for Modeling Emotion'.

Hudlicka, E. (2008), Modeling the mechanisms of emotion effects on cognition., *in* 'AAAI Fall Symposium: Biologically Inspired Cognitive Architectures', Vol. FS-08-04 of *AAAI Technical Report*, AAAI, pp. 82–86.

Hudlicka, E. (2009), Affective game engines: Motivation and requirements, *in* 'Proceedings of the 4th International Conference on Foundations of Digital Games', FDG '09, ACM, New York, NY, USA, pp. 299–306.

Hudlicka, E. (2011), 'Affective gaming in education, training and therapy: Motivation, requirements, techniques', *Handbook of Research on Improving Learning and Motivation through Educational Games: Multidisciplinary Approaches* pp. 482–511.

IVA (2016), 'Intelligent virtual agents international conference'.
URL: *http://dblp.uni-trier.de/db/conf/iva/index.html*

Jack, R. E., Garrod, O. G., Yu, H., Caldara, R. & Schyns, P. G. (2012), 'Facial expressions of emotion are not culturally universal', *Proceedings of the National Academy of Sciences* **109**(19), 7241–7244.

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T. & Walton, A. (2008), 'Measuring and defining the experience of immersion in games', *International journal of human-computer studies* **66**(9), 641–661.

Jiang, B., Valstar, M., Martinez, B. & Pantic, M. (2014), 'A dynamic appearance descriptor approach to facial actions temporal modeling', *Cybernetics, IEEE Transactions on* **44**(2), 161–174.

Johannes Wagner, E. A. & Jung, F. (2009), Smart sensor integration: A framework for multimodal emotion recognition in real-time, *in* 'Affective Computing and Intelligent Interaction (ACII 2009)', IEEE.

John, O. P., Naumann, L. P. & Soto, C. J. (2008), 'Paradigm shift to the integrative big five trait taxonomy', *Handbook of personality: Theory and research* **3**, 114–158.

Jones, H., Cholet, M., Ochs, M., Sabouret, N. & Pelachaud, C. (2014), Expressing social attitudes in virtual agents for social coaching, *in* 'Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems', AAMAS '14, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, pp. 1409–1410.
URL: *http://dl.acm.org/citation.cfm?id=2617388.2617497*

Jung, B. & Kopp, S. (2003), Flurmax: An interactive virtual agent for entertaining visitors in a hallway, *in* 'Intelligent virtual agents', Springer, pp. 23–26.

Karimi, S. & Kangavari, M. R. (2012), 'A computational model of personality', *Procedia-Social and Behavioral Sciences* **32**, 184–196.

Kelley, J. F. (1984), 'An iterative design methodology for user-friendly natural language office information applications', *ACM Trans. Inf. Syst.* **2**(1), 26–41.

Kemper, T. D. (1973), 'The fundamental dimensions of social relationship: A theoretical statement', *Acta Sociologica* **16**(1), 41–60.

Kendon, A. (1990), *Conducting Interaction: Patterns of Behavior in Focused Encounters*, Cambridge Music Handbooks, Cambridge University Press.
URL: *http://books.google.co.uk/books?id=m9qsQgAACAAJ*

Knapp, M., Hall, J. & Horgan, T. (2013), *Nonverbal communication in human interaction*, Cengage Learning.

Kopp, S., Gesellensetter, L., Krmer, N. C. & Wachsmuth, I. (2005), A conversational agent as museum guide - design and evaluation of a real-world application., *in* T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier & T. Rist, eds, 'IVA', Vol. 3661 of *Lecture Notes in Computer Science*, Springer, pp. 329–343.

Kopp, S., Krenn, B., Marsella, S., Marshall, A. N., Pelachaud, C., Pirker, H., Thrisson, K. R. & Vilhjlmsson, H. (2006), Towards a common framework for multimodal generation: The behavior markup language, *in* 'INTERNATIONAL CONFERENCE ON INTELLIGENT VIRTUAL AGENTS', pp. 21–23.

Kopp, S., Krenn, B., Marsella, S. & et al. (2006), Towards a common framework for multimodal generation: The behavior markup language, *in* 'Int'l Conf. Intelligent Virtual Agents', pp. 21–23.

Krantz, J. (2012), *Experiencing Sensation and Perception*, Pearson Education, Limited.

Kshirsagar, S. & Magnenat-Thalmann, N. (2002), A multilayer personality model, *in* 'Proceedings of the 2Nd International Symposium on Smart Graphics', SMARTGRAPH '02, ACM, New York, NY, USA, pp. 107–115.

Laird, J. E. & Duchi, J. C. (2000), 'Creating human-like synthetic characters with multiple skill levels: A case study using the soar quakebot', *Ann Arbor* **1001**, 48109–2110.

Laird, J. E., Newell, A. & Rosenbloom, P. S. (1987), 'Soar: An architecture for general intelligence', *Artificial Intelligence* **33**(1), 1 – 64.

Lazarus, R. S. (1982), 'Thoughts on the relations between emotion and cognition.', *American psychologist* **37**(9), 1019.

Lazarus, R. S. (1991), *Emotion and Adaptation*, Oxford University Press, USA.

Lazzaro, N. (2004), 'Why we play games: Four keys to more emotion without story'.

Le, Q. A. & Pelachaud, C. (2012), *Generating Co-speech Gestures for the Humanoid Robot NAO through BML*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 228–237.

Lebiere, C. & Anderson, J. R. (2008), 'A connectionist implementation of the act-r production system'.

Lee, M. S. & Heeter, C. (2015), 'Cognitive intervention and reconciliation: Npc believability in single-player rpgs', *International Journal of Role-Playing* (5), 47–65.

Li, B., Thakkar, M., Wang, Y. & Riedl, M. O. (2014), 'Data-driven alibi story telling for social believability', *Social Believability in Games* .

Li, T., Qiu, Y., Yue, P. & Zhong, G. (2007), Exploiting model of personality and emotion of learning companion agent, *in* 'Computer Systems and Applications, 2007. AICCSA '07. IEEE/ACS International Conference on', pp. 860–865.

Lövheim, H. (2012), 'A new three-dimensional model for emotions and monoamine neurotransmitters', *Medical Hypotheses* **78**(2), 341 – 348.

Magnenat-Thalmann, N., Bonanni, U. & Volino, P. (2009), Physical behavior of deformable hair and clothes: What is common?, *in* 'Proc. of the 11th IEEE International conference on Computer-Aided Design and Computer Graphics (CAD/Graphics 2009)', IEEE Comp. Soc., pp. 12–18.

Magnenat-Thalmann, N., Hadap, S. & Kalra, P. (2000), State of the art in hair simulation.

Magnenat-Thalmann, N. & Thalmann, D. (1982), Dreamflight: a fictional film realized by 3d computer animation, *in* 'Proc. ONLINE Conf., London, UK', pp. 353–367.

Magnenat-Thalmann, N. & Thalmann, D. (1987), 'The direction of synthetic actors in the film rendez-vous a montreal', *IEEE Comput. Graph. Appl.* **7**(12), 9–19.

Magnenat Thalmann, N. & Thalmann, D. (1995), 'Digital actors for interactive television', *Proceedings of the IEEE* **83**(7), 1022–1031.

Magnenat-Thalmann, N. & Thalmann, D. (2004), *An Overview of Virtual Humans*, John Wiley, pp. 1–25.

Magnenat-Thalmann, N. & Thalmann, D. (2005), 'Virtual humans: thirty years of research, what next?', *The Visual Computer* **21**(12), 997–1015.

Magnenat-Thalmann, N., Thalmann, D. & Beland, S. (1986), The integration of particles and polygons rendering using a z-buffer algorithm, *in* 'Proc. of the European Conference and Exhibition, Lisbon, Portugal (Eurographics'86)'.

Magnenat-Thalmann, N., Volino, P., Bonanni, U., Summers, I. R., Bergamasco, M., Salsedo, F. & Wolter, F. E. (2007), 'From physics-based simulation to the touching of textiles: The haptex project', *The International Journal of Virtual Reality* **6**(3), 35–44.

Mandler, G. (1984), *Mind and Body: Psychology of Emotion and Stress*, W W Norton and Co Inc.

Marsella, S., Gratch, J. & Petta, P. (2010), 'Computational models of emotion', *A Blueprint for Affective Computing-A sourcebook and manual* pp. 21–46.

Marston, W. M. (2013), *Emotions of normal people*, Vol. 158, Routledge.

Maslow, A. H. (1943), 'A theory of human motivation.', *Psychological review* **50**(4), 370.

McCrae, R. R. & John, O. P. (1992), 'An introduction to the five-factor model and its applications', *Journal of personality* **60**(2), 175–215.

McKeown, G., Valstar, M., Cowie, R., Pantic, M. & Schröder, M. (2012), 'The semaine database: Annotated multimodal records of emotionally coloured conversations between a person and a limited agent', *T. Affective Computing* **3**(1), 5–17.

McRorie, M., Sneddon, I., McKeown, G., Bevacqua, E., de Sevin, E. & Pelachaud, C. (2012), 'Building and Evaluating Personality in Virtual Agents', *Transactions on Affective Computing* .

Mehrabian, A. (1996), 'Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament', *Current Psychology* **14**(4), 261–292.

Mehrabian, A. & Russell, J. A. (1973), 'A measure of arousal seeking tendency.', *Environment and Behavior* .

Miikkulainen, R. (2007), Creating intelligent agents in games, *in* 'Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2006 Symposium', National Academies Press, p. 15.

Molet, T., Aubel, A., Capin, T. K., Carion, S., Lee, E., Magnenat-Thalmann, N., Noser, H., Pandzic, I. S., Sannier, G. & Thalmann, D. (1999), 'Anyone for tennis?', *Presence* **8**(2), 140–156.

Mui, L., Mohtashemi, M. & Halberstadt, A. (2002), A computational model of trust and reputation, *in* 'System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on', IEEE, pp. 2431–2439.

Murphy, G. & Kovach, J. (1972), *Historical introduction to modern psychology*, Harcourt Brace Jovanovich.

Myers, I. B., McCaulley, M. H., Quenk, N. L. & Hammer, A. L. (1998), *MBTI manual: A guide to the development and use of the Myers-Briggs Type Indicator*, Vol. 3, Consulting Psychologists Press Palo Alto, CA.

Newell, A. (1994), *Unified Theories of Cognition*, William James Lectures, Harvard University Press.

URL: *http://books.google.co.uk/books?id=1lbY14DmV2cC*

Nguyen, N. & Wachsmuth, I. (2009), Modeling peripersonal action space for virtual humans using touch and proprioception, *in* Z. Ruttkay, M. Kipp, A. Nijholt & H. Vilhjlmsson, eds, 'Intelligent Virtual Agents', Vol. 5773 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 63–75.

Nguyen, N. & Wachsmuth, I. (2011), From body space to interaction space: Modeling spatial cooperation for virtual humans, *in* 'The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 3', AAMAS '11, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, pp. 1047–1054.

Nguyen, N., Wachsmuth, I. & Kopp, S. (2007), Touch perception and emotional appraisal for a virtual agent, *in* 'Proceedings Workshop Emotion and Computing-Current Research and Future Impact'.

Norman, W. T. (1963), 'Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings.', *The Journal of Abnormal and Social Psychology* **66**(6), 574.

Ochs, M., Sabouret, N. & Corruble, V. (2009), 'Simulation of the dynamics of nonplayer characters' emotions and social relations in games', *Computational Intelligence and AI in Games, IEEE Transactions on* **1**(4), 281–297.

Ortony, A. (1991), 'Memories, thoughts, and emotions: Essays in honor of george mandler, chapter value and emotion'.

Ortony, A., Clore, G. & Collins, A. (1988), *Cognitive Structure of Emotions*, Cambridge University Press.

Pecune, F., Ochs, M., Marsella, S. & Pelachaud, C. (2016), Socrates: From social relation to attitude expressions, *in* 'Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems', AAMAS '16, International Foundation for

Autonomous Agents and Multiagent Systems, Richland, SC, pp. 921–930.
URL: *http://dl.acm.org/citation.cfm?id=2937029.2937060*

Pecune, F., Ochs, M. & Pelachaud, C. (2013), A formal model of social relations for artificial companions, *in* 'EUMAS'.

Pelachaud, C. & Ochs, M. (2015), Towards a socially adaptive virtual agent, *in* 'Intelligent Virtual Agents: 15th International Conference, IVA 2015, Delft, The Netherlands, August 26-28, 2015, Proceedings', Vol. 9238, Springer, p. 3.

Pentland, A. S. (2005), 'Socially aware computation and communication', *Computer* **38**(3), 33–40.

Pereira, F. C. & Ebrahimi, T. (2002), *The MPEG-4 Book*, Prentice Hall PTR, Upper Saddle River, NJ, USA.

Peters, C., Pelachaud, C., Bevacqua, E., Ochs, M., Chafai, N. E. & Mancini, M. (2006), Social capabilities for autonomous virtual characters, *in* 'International Digital Games Conference', pp. 37–48.

Picard, R. W. (1997), *Affective Computing*, MIT Press, Cambridge, MA, USA.

Plutchik, R. (1980), *Emotion: A Psychoevolutionary Synthesis*, Harper and Row.
URL: *http://books.google.co.uk/books?id=G5t9AAAAMAAJ*

Prada, R., Raimundo, G., Baptista, M., Dimas, J., Santos, P. A., Martinho, C., Pea, J. & Ribeiro, L. L. (2012), The role of social identity, rationality and anticipation in believable agents, *in* 'Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)', International Foundation for Autonomous Agents and Multiagent Systems, pp. 1175–1176.

Preda, M. & Preteux, F. (2002), Critic review on mpeg-4 face and body animation, *in* 'Image Processing. 2002. Proceedings. 2002 International Conference on', Vol. 3, pp. 505–508 vol.3.

Preda, M. & Preteux, F. (2004), 'Virtual character within mpeg-4 animation framework extension', *IEEE Transactions on Circuits and Systems for Video Technology* **14**(7), 975–988.

Preethi Narasimhan, K. & White, G. (2013), An agent-based analyses of f-formations, *in* Y. Demazeau, T. Ishida, J. Corchado & J. Bajo, eds, 'Advances on Practical Applications of Agents and Multi-Agent Systems', Vol. 7879 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 239–250.

Querrec, R. & Chevaillier, P. (2001), *Virtual Storytelling for Training: An Application to Fire Fighting in Industrial Environment*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 201–204.

Rammstedt, B. & John, O. P. (2007), 'Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german', *Journal of research in Personality* **41**(1), 203–212.

Rao, A. S. & Georgeff, M. P. (1995), Bdi agents: From theory to practice, *in* 'ICMAS-95', pp. 312–319.

Raskin, V. (2012), *Semantic mechanisms of humor*, Vol. 24, Springer Science & Business Media.

Reisenzein, R., Hudlicka, E., Dastani, M., Gratch, J., Hindriks, K., Lorini, E. & Meyer, J.-J. (2013), 'Computational modeling of emotion: Toward improving the inter- and intradisciplinary exchange', *Affective Computing, IEEE Transactions on* **4**(3), 246–266.

Richmond, V., McCroskey, J. & Payne, S. (1991), *Nonverbal behavior in interpersonal relations*, Prentice Hall.
URL: *http://books.google.co.uk/books?id=aFodAQAAMAAJ*

Riedl, M. O. & Stern, A. (2006), Believable agents and intelligent story adaptation for interactive storytelling, *in* 'International Conference on Technologies for Interactive Digital Storytelling and Entertainment', Springer, pp. 1–12.

Ringeval, F., Schuller, B., Michel, V., Jaiswal, S., Marchi, E., Lalanne, D., Cowie, R. & Pantic, M. (2015), The av+ec 2015 multimodal affect recognition challenge: Bridging across audio, video, and physiological data, *in* 'The Audio/Visual Emotion Challenge and Workshop(AV+EC 2015), The ACM Multimedia Conference for 2015'.

Rizzo, A. S., Shilling, R., Forbell, E., Scherer, S., Gratch, J. & Morency, L.-P. (2015), Autonomous Virtual Human Agents for Healthcare Information Support and Clinical Interviewing, Elsevier, Inc., Philadelphia, PA, pp. 53–79.

Rizzo, A. & Talbot, T. (2016), Virtual Reality Standardized Patients for Clinical Training, *in* 'The Digital Patient', John Wiley & Sons, Inc, Hoboken, NJ, pp. 255–272.

Robinson, S., Traum, D., Ittycheriah, M. & Henderer, J. (2008), What would you ask a Conversational Agent? Observations of Human-Agent Dialogues in a Museum Setting, *in* 'International Conference on Language Resources and Evaluation (LREC)'.

Roseman, I. J. (1996), 'Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory', *Cognition and Emotion* **10**(3), 241–278.

Rosenbloom, P., Laird, J. & Newell, A. (1993), *The Soar papers: research on integrated intelligence*, number v. 2 *in* 'Artificial intelligence', MIT Press.

Rosis, F. D., Pelachaud, C., Poggi, I., Carofiglio, V. & Carolis, B. D. (2003), 'From gretas mind to her face: Modelling the dynamics of affective states in a conversational embodied agent', *International Journal of Human-Computer Studies* **59**, 81–118.

Rousseau, D. et al. (1996), Personality in computer characters, *in* 'Proceedings of the 1996 AAAI Workshop on Entertainment and AI/A-Life', pp. 38–43.

Russell, J. A. (1980), 'A circumplex model of affect.', *Journal of personality and social psychology* **39**(6), 1161.

Rusting, C. L. & Larsen, R. J. (1997), 'Extraversion, neuroticism, and susceptibility to positive and negative affect: A test of two theoretical models', *Personality and Individual Differences* **22**(5), 607–612.

Salber, D. & Coutaz, J. (1993), Applying the wizard of oz technique to the study of multimodal systems, *in* 'In Human Computer Interaction, 3rd International Conference EWHCI93', Springer Verlag, pp. 219–230.

Saucier, G. & Goldberg, L. R. (1996), The language of personality: Lexical perspectives on the five-factor model, *in* J. S. Wiggins, ed., 'The five-factor model of personality: Theoretical perspectives', New York: Guilford, pp. 21–50.

Schell, J. (2014), *The Art of Game Design: A book of lenses*, CRC Press.

Scherer, S., Marsella, S. C., Stratou, G., Xu, Y., Morbini, F., Egan, A., Rizzo, A. & Morency, L.-P. (2012), Perception markup language: Towards a standardized representation of perceived nonverbal behaviors, *in* 'The 12th International Conference on Intelligent Virtual Agents (IVA)', Santa Cruz, CA.

Schröder, M. (2010), 'The SEMAINE API: Towards a standards-based framework for building emotion-oriented systems', *Advances in Human-Computer Interaction* **2010**.

Schroder, M., Bevacqua, E., Cowie, R., Eyben, F., Gunes, H., Heylen, D., ter Maat, M., McKeown, G., Pammi, S., Pantic, M., Pelachaud, C., Schuller, B., de Sevin, E., Valstar, M. & Wollmer, M. (2012), 'Building autonomous sensitive artificial listeners', *Affective Computing, IEEE Transactions on* **3**(2), 165–183.

Schuller, B., Batliner, A., Steidl, S. & Seppi, D. (2011), 'Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge', *Speech Communication* **53**(9), 1062–1087.

Schuller, B., Eyben, F., Gunes, H., Pantic, M., Schröder, M., Valstar, M. & Wöllmer, M. (n.d.), 'Final face and voice feature extraction component with incremental, (near) real-time processing'.

Schuller, B., Rigoll, G. & Lang, M. (2004), Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture, *in* 'Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on', Vol. 1, IEEE, pp. I–577.

Schuller, B., Steidl, S., Batliner, A., Schiel, F. & Krajewski, J. (2014), 'Introduction to the special issue on broadening the view on speaker analysis', *Computer Speech and Language* **28**(2), 343–345.

Schutz, W. C. (1958), 'Firo: A three-dimensional theory of interpersonal behavior.'.

Shahid, S., Krahmer, E. & Swerts, M. (2011), Child-robot interaction: Playing alone or together?, *in* 'CHI '11 Extended Abstracts on Human Factors in Computing Systems', CHI EA '11, ACM, New York, NY, USA, pp. 1399–1404.

Shahid, S., Krahmer, E. & Swerts, M. (2012), 'Video-mediated and co-present gameplay: Effects of mutual gaze on game experience, expressiveness and perceived social presence', *Interact. Comput.* **24**(4), 292–305.

Shapiro, A. (2011), Building a character animation system, *in* J. Allbeck & P. Faloutsos, eds, 'Motion in Games', Vol. 7060 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 98–109.

Shapiro, S. S. & Wilk, M. B. (1965), 'An analysis of variance test for normality (complete samples)', *Biometrika* pp. 591–611.

Shattuck, H. D. (1988), 'Emotion and behaviour'.

Sidner, C., Rich, C., Shayganfar, M., Bickmore, T., Ring, L. & Zhang, Z. (2015), A robotic companion for social support of isolated older adults, *in* 'Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts', HRI'15 Extended Abstracts, ACM, New York, NY, USA, pp. 289–289.

Smith, C. A. & Ellsworth, P. C. (1985), 'Patterns of cognitive appraisal in emotion', **48**, 813–838+.

URL: *http://search.ebscohost.com/login.aspx?direct=true&db=pdh&AN=psp-48-4-813&site=ehost-live*

Smith, C. A. & Lazarus, R. S. (1993), 'Appraisal components, core relational themes, and the emotions', *Cognition & Emotion* **7**(3-4), 233–269.

Sollenberger, D. & Singh, M. (2012), 'Koko: An architecture for affect-aware games', *Autonomous Agents and Multi-Agent Systems* **24**(2), 255–286.

Steunebrink, B. R., Dastani, M. & Meyer, J.-J. C. (2008), A formal model of emotions: Integrating qualitative and quantitative aspects, *in* 'Proceedings of the 2008 Conference on ECAI 2008: 18th European Conference on AI', IOS Press, Amsterdam, The Netherlands, The Netherlands, pp. 256–260.

Stratou, G., Morency, L.-P., DeVault, D., Hartholt, A., Fast, E., Lhommet, M., Lucas, G., Morbini, F., Georgila, K., Scherer, S., Gratch, Jonathan, Marcella Stacy, Traum, David & Rizzo, Albert (2015), A Demonstration of the Perception System in SimSensei, a Virtual Human Application for Healthcare Interviews, *in* 'Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on', IEEE, Xi'an, China, pp. 787–789.

Strongman, K. (2003), *The psychology of emotion: from everyday life to theory*, Wiley.

Sun, R. (2006), 'The clarion cognitive architecture: Extending cognitive modeling to social simulation', *Cognition and multi-agent interaction* pp. 79–99.

Sun, R. (2007), 'The importance of cognitive architectures: An analysis based on clarion', *Journal of Experimental and Theoretical Artificial Intelligence* .

Sun, R. (2016), *Anatomy of the Mind: Exploring Psychological Mechanisms and Processes with the Clarion Cognitive Architecture*, Oxford University Press.

Sutcliffe, A. & Wang, D. (2012), 'Computational modelling of trust and social relationships', *Journal of Artificial Societies and Social Simulation* **15**(1), 3.

Swartout, W., Traum, D., Artstein, R., Noren, D., Debevec, P., Bronnenkant, K., Williams, J., Leuski, A., Narayanan, S., Piepol, D. et al. (2010), Ada and grace: Toward realistic and engaging virtual museum guides, *in* 'International Conference on Intelligent Virtual Agents', Springer, pp. 286–300.

ter Maat, M. & Heylen, D. (2011), 'IVA: 10th international conference, IVA 2011, reykjavik, september 15-17, 2011', pp. 470–472.

Thalmann, D. (1996), Physical, behavioral, and sensor-based animation, *in* 'Proc. Graphicon 96,', number VRLAB-CONF-2007-102, pp. 214–221.

Thiebaux, M., Marshall, A., Marsella, S. C. & Kallmann, M. (2008), SmartBody: Behavior realization for embodied conversational agents, *in* 'International Conference on Autonomous Agents and Multiagent Systems (AAMAS)', Estoril, Portugal.

Thomas, F. & Johnson, O. (1984), *Disney Animation: The Illusion of Life*, Abbeville Press.

Traum, D. R. & Larsson, S. (2003), The information state approach to dialogue management, *in* 'Current and new directions in discourse and dialogue', Springer, pp. 325–353.

Turing, A. M. (1950), 'Computing machinery and intelligence', *Mind* **59**(236), pp. 433–460.

Urbain, J., Niewiadomski, R., Mancini, M. & et al. (2013), Multimodal analysis of laughter for an interactive system, *in* 'Intelligent Technologies for Interactive Entertainment', Vol. 124, pp. 183–192.

Vala, M., Sequeira, P., Paiva, A. & Aylett, R. (2007), Fearnot! demo: a virtual environment with synthetic characters to help bullying, *in* '6th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2007)', Vol. 271, IFAAMAS.

Valstar, M. (2015), Automatic facial expression analysis, *in* 'Understanding Facial Expressions in Communication', Springer India, pp. 143–172.

Valstar, M. F. (2008), Timing is everything : a spatio-temporal approach to the analysis of facial actions, PhD thesis, Imperial Collage.

Valstar, M., Jiang, B., Mehu, M., Pantic, M. & Scherer, K. (2011), The first facial expression recognition and analysis challenge, *in* 'Automatic Face Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on', pp. 921–926.

Valstar, M. & Pantic, M. (2006), Fully automatic facial action unit detection and temporal analysis, *in* 'Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on', IEEE, pp. 149–149.

Vardoulakis, L. P., Ring, L., Barry, B., Sidner, C. L. & Bickmore, T. (2012), Designing relational agents as long term social companions for older adults, *in* 'Proceedings of the 12th International Conference on Intelligent Virtual Agents', Springer-Verlag, Berlin, Heidelberg, pp. 289–302.

Velsquez, J. D. (1997), Modeling emotions and other motivations in synthetic agents, *in* 'IN: PROCEEDINGS OF AAAI97'.

Verhoef, T., Lisetti, C., Barreto, A., Ortega, F., Zant, T. & Cnossen, F. (2009), Bio-sensing for emotional characterization without word labels, *in* J. Jacko, ed., 'Human-Computer Interaction. Ambient, Ubiquitous and Intelligent Interaction', Vol. 5612 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 693–702.

Vernon, D., Metta, G. & Sandini, G. (2007), 'A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents', *Evolutionary Computation, IEEE Transactions on* **11**(2), 151–180.

Vilhjalmsson, H., Badler, N., Johnson, L., Kopp, S., Krenn, B., Marsella, S., Marshall, A. N., Pelachaud, C., Pirker, H. & Thorisson, K. R. (2016), 'Saiba framework'.
URL: *http://www.mindmakers.org/projects/saiba/wiki*

Vilhjalmsson, H., Cantelmo, N., Cassell, J., Chafai, N. E., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A. N., Pelachaud, C., Ruttkay, Z., Thrisson, K. R., Welbergen, H. V. & Werf, R. J. V. D. (2007), 'The behaviour markup language: recent developments and challenges', *LECTURE NOTES IN ARTIFICIAL INTELLIGENCE* **4722**, 99–11.

Vilhjalmsson, H., Cantelmo, N.and Cassell, J. & et al. (2007), 'The behaviour markup language: recent developments and challenges', *Lecture Notes in Artificial Intelligence* **4722**, 99–11.

Villon, O. & Lisetti, C. (2006), Toward building adaptive users psycho-physiological maps of emotions using bio-sensors, *in* 'In Proceedings of KI06 26th German Annual Conference in Artificial Intelligence'.

Vinciarelli, A., Pantic, M. & Bourlard, H. (2009), 'Social signal processing: Survey of an emerging domain', *Image Vision Comput.* **27**(12), 1743–1759.

Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F. & Schroeder, M. (2012), 'Bridging the gap between social animal and unsocial machine: A survey of social signal processing', *Affective Computing, IEEE Transactions on* **3**(1), 69–87.

Volino, P. & Magnenat-Thalmann, N. (2001), Comparing efficiency of integration methods for cloth simulation, *in* 'Computer Graphics International 2001', CGI '01, IEEE Computer Society, Washington, DC, USA, pp. 265–274.

Volino, P., Magnenat-Thalmann, N. & Faure, F. (2009), 'A simple approach to nonlinear tensile stiffness for accurate cloth simulation', *ACM Transactions on Graphics* pp. 105–116.

Wagner, J., Lingenfelser, F., Baur, T., Damian, I., Kistler, F. & Andrè, E. (2013), The social signal interpretation (SSI) framework: Multimodal signal processing and recognition in real-time, *in* 'Proceedings of the 21st ACM International Conference on Multimedia', pp. 831–834.

Wagner, J., Lingenfelser, F., Bee, N. & Andr, E. (2011), 'Social signal interpretation (ssi) - a framework for real-time sensing of affective and social signals.', *KI* **25**(3), 251–256.

Walker, W., Lamere, P., Kwok, P., Raj, B., Singh, R., Gouvea, E., Wolf, P. & Woelfel, J. (2004), Sphinx-4: A flexible open source framework for speech recognition, Technical report, Mountain View, CA, USA.

Wallace, R. (2003), 'The elements of aiml style', *Alice AI Foundation* .

Watson, D. & Clark, L. A. (1992), 'On traits and temperament: General and specific factors of emotional experience and their relation to the five-factor model', *Journal of personality* **60**(2), 441–476.

Wooldridge, M. & Jennings, N. R. (1995), 'Intelligent agents: theory and practice', *The Knowledge Engineering Review* **10**, 115–152.

Yannakakis, G. N. & Paiva, A. (2014), 'Emotion in games', *Handbook on affective computing* pp. 459–471.

Yuan, X. & Chee, Y. S. (2003), Embodied tour guide in an interactive virtual art gallery, *in* 'Proceedings. 2003 International Conference on Cyberworlds', pp. 432–439.